

20 November 1980

Deciding where to put nuclear plants

Even the most embittered opponents of nuclear power must be surprised at the good luck that has come their way in the past few months. First, there was the accident at Three Mile Island in March 1979, the chief casualty of which was the reputation of the US Nuclear Regulatory Agency. In the wake of the report of the Presidential Commission of Inquiry, it was inevitable that the regulatory commission would have to be reorganized and that its chairman, Dr Joseph M. Hendrie, would eventually be required to practise his talents elsewhere. In due course, President Carter nominated a new chairman (Professor Al Carnesale), but so soon before the election that his appointment to the job was legitimately blocked by the Republicans in Congress and will now, presumably, fall by the wayside. In the interregnum, the commission has been behaving with accustomed but quite alarming timorousness. The second reactor at Three Mile Island, entirely undamaged and now more fully understood than ever, is still out of service. The granting of operating licences for plants already built has proceeded at a snail's pace. Then, almost as if to ensure that no utility will be so rash as to order a nuclear plant for some time to come, the commission embarked at the end of July on a series of inquiries aimed at defining the criteria to be met by the sites proposed for nuclear plants. The outcome of this inquiry, known technically as a rulemaking procedure and thus designed to yield rules applicable throughout the United States, is naturally unforeseeable. Several years, perhaps five, may pass before putative builders of nuclear plants know where they stand. If they are wise, they will in the meantime put their capital into growing trees. That way, they will at least have something to burn.

Elsewhere than in the United States, stasis is less complete but of longer standing. In Britain, for example, it now seems likely that the Central Electricity Generating Board will be allowed to order one new nuclear plant in each of the next two years, but that 1982 will be largely given over to a public inquiry about the proposal to build a single pressurized water reactor of Westinghouse design. (The Three Mile Island reactors use pressurized water, but were built by Babcock.) As yet, the terms of reference of the inquiry (promised by Mr Tony Benn, previously Mr Anthony Wedgwood Benn, when a minister in the previous government) have not been decided. It remains to be seen whether the inquiry will be organized around a site-specific proposal or, alternatively, drawn in such a way that pressurized water reactors as such will be pronounced usable or otherwise. But only after that inquiry is complete will the same group of people embark on the more contentious job of preparing for an inquiry into the safety of full-sized fast reactors, now on the calendar for 1984.

Both the opponents of the nuclear construction industry and those who think it prudent to build more of them must surely agree that these are not ways in which decisions should be made about the future development of nuclear power. Whatever the rights and wrongs of the safety argument, procrastination cannot substitute for rational decision, although it may be responsible for decision by default. Since the problem of winning public acceptance, or at least of creating a sense that justice has been done, applies to other kinds of technological enterprises than the construction of nuclear power stations, it is obviously important that better mechanisms should be devised.

The regulation of nuclear power plants in the United States has been a controversial issue for the best part of twenty years. To begin with, the Atomic Energy Commission was both the sponsor of nuclear power and the licensing authority. Understandably,

and rightly, the two functions were separated a decade ago. Again rightly, the Nuclear Regulatory Commission is an independent body whose members are appointed by the president of the day for fixed terms, and who are collectively responsible to Congress and not the Administration for the proper regulation of nuclear safety. Like other regulatory agencies, the Federal Communications Commission for example, the Nuclear Regulatory Commission enjoys a kind of delegated authority to make rules which elsewhere would have the status of laws. It is empowered to make generic rules applying, say, to all reactors of a specific type, and also particular rules, as on the licensing for operation of a particular reactor. Other government agencies are powerless to influence decisions by the commissions, but those who dissent can appeal to the courts. Such appeals have been one of the most frequent means of impeding the construction of nuclear plants in the United States.

The British system is at the other end of the regulatory spectrum. Broadly speaking, all new industrial developments are subject to planning legislation. Applications for planning consent are first submitted to the lowest planning authority in the hierarchy, a county council for example. Necessarily, they are site-specific. One applicant may want to build an abattoir at the end of some village high street, another (invariably a nationalized industry) a nuclear power station on an estuary. The planning legislation provides that the Secretary of State at the Department of the Environment may order a public inquiry if either the proposers or the opponents of some innovation are aggrieved, or if he and his officials consider that the issue has wider importance. The planning legislation also allows for the appointment of commissions of inquiry to determine generic questions — are motorways, or nuclear power stations, or fast reactors, really necessary, for example? But this procedure has not so far been used, no doubt for fear of creating precedents.

Both the British and the American systems are intended to serve similar ends — to allow objectors, especially disappointed objectors, a proper sense that their objections have been fully heard, but also to give the public at large a sense that important questions have been properly examined. That, at least, is the theory. The reality is usually less than satisfying. Private objectors are usually less well equipped, at least financially, to put their case to government inspectors. In the United States, the Internal Revenue Service allows charitable foundations to contribute to the cost of putting objections to a public hearing. In Britain, where this is not permitted, the government might think it worthwhile making some funds available for objectors to important schemes. (Inspectors at present have a right to penalize frivolous objectors with an order to pay costs, but that is hardly the same thing.) A more serious difficulty under British planning law is that objectors to a novel scheme have only restricted rights to argue that the objective of a new plant might be better achieved by siting it somewhere else.

These, however, are relatively minor difficulties. The most serious objection to both the British and the American systems for deciding whether and, if so, where nuclear power stations should be built is that both of them are inevitably subject to political forces in the broadest sense, and are thus far from being objective. In Britain, the more fuss that can be made about a proposal to build a nuclear power station, the greater the chance that there will be a public inquiry. All nuclear plants so far built have been brought into the world with ceremonies of this kind. No minister with an eye to re-election would think otherwise. In the United



States, similar pressures affect the Nuclear Regulatory Commission, whose reputation would quickly sink still further if it were often defeated in the courts on the grounds of not having been scrupulous about due process. And at inquiries of both kinds, there is nothing to prevent the most local objectors from raising the biggest questions that come to mind. Does the country need nuclear power anyway? Is this kind of reactor safe? Why should Blanksville or Dashby be the first prisoner of the plutonium economy, as it is called? In the United States, these pressures have led to a quite absurd proliferation of the subjects on which the Nuclear Regulatory Commission is holding public hearings. The inevitable result is more prevarication, more delay.

As things are, the only seemly escape from the dilemma is to devise a means of limiting the scope of public hearings without limiting the right of objectors to object. On the construction of

nuclear plants, it is pointless that public hearings should rehash questions of whether nuclear power as such is necessary, or economic, or likely to be overtaken by solar power. Both Britain and the United States have elected governments which reckon to include such questions in the fields of their competence. Similarly, there are questions about the relative (not absolute) safety of different reactor types which are generic in character, and which could (and should) be argued out once and for all. Questions of siting, on the other hand, are essentially local or regional matters (which is not to imply that they should not be disputed nationally). But it should be open to objectors to argue that some other site would be more suitable. The British log-jam of nuclear inquiries will not be broken until 1984. Is there a chance that Mr Reagan will be able to move more quickly in the United States when January has come?

Scope for European collaboration?

The famous conjuring trick in which the Cheshire Cat was made to disappear but to leave its smile behind is no doubt the envy of the professionals who still work the music-halls. The European Science Foundation, which held its annual assembly last week in Strasbourg (see page 204), is trying to do much the same. Its members are research councils and learned academies from most Western European nations, all of them supported financially by their national governments. Their membership of the foundation is voluntary, as are their contributions to the special research projects which are from time to time devised. It is just as if they belonged to a club for organizations like themselves and, like the members of other respectable clubs, they spend some time each year considering what the fees for membership should be. The trick, which has worked well in the past six years, lies in the foundation's insistence that it is strictly a non-governmental organization, free (if its members agree) to criticize developments at which its members' paymaster-governments may have connived. This is the spirit in which the European Space Agency has been taken to task in the past two years for the inadequacy of its plans for launching scientific satellites. On other occasions, the foundation assumes the freedom to put pressure on the same governments to act in some specific way — to build a synchrotron radiation source, for example. The hapless governments see the smile (or the grimace) but appear not to recognize the body.

That, fortunately, has been the foundation's experience so far. Part of the explanation is that it has been run on a kind of shoestring. The members' contributions to the foundation are such modest proportions of the national public funds they receive that even the most hyper-touchy governments would feel foolish if they reacted against the foundation's members' modest show of independence. But will it be the same if the foundation grows, becoming less inconspicuous in the process? That, inevitably, must be one of the concerns of those who now think that the time has come for modest growth. The time has come to ask what the foundation is for, and what it might be for. For the past few months, a domestic think-tank has been brooding on the same two questions.

The bread and butter of the foundation's business is easily justified. Plainly it is useful that there should be an umbrella under which organizations with a responsibility for the support of research and scholarships can meet from time to time, even if they talk only about administrative matters such as how to adjust research grants for inflation. In principle, it is also plainly beneficial that there should be some mechanism for tackling research problems of common European concern. European taxonomy is an obvious candidate. So too was the project to mount a coordinated study of the problems of preserving mediaeval stained glass, but that fell foul of the chauvinistic quarrelsomeness of the experts in the field. Unexpectedly, but not surprisingly, most enthusiasm seems to have been generated by the common interest of European scholars in external problems — the literature of China and Byzantine history, for example.

Collaborative research programmes within Europe itself seem more easily arranged in the social than the hard sciences — partly because of the costs involved but also because international comparisons come more naturally to mind in the social sciences; in the hard sciences, inevitably, everything is everywhere the same. Yet there is obviously plenty of scope for further collaborative research. The geophysicists have for example an ambition to mount a seismic traverse of Europe from north to south, reaching into North Africa, that could probably be mounted within the necessarily limited time-scale of *ad hoc* voluntary contributions by the member organizations. The plan to build a source of synchrotron radiation would, by contrast, require a continuing commitment of funds, and thus endanger the loyalty of some members or the compliance of their governments. The strategy the foundation has pursued — to interest governments in the idea — is thus the only practicable way of seeing such a machine built.

So where does the foundation go from here? For the time being, there are grounds for being cautious. Inevitably, in only six years, the foundation has comparatively little to show for the trouble it has been taking. Ultimately — and there should not be long to wait — it will be judged by the quality of its collaborative research. If the outcome is worthwhile, the result should be a rash of other projects crying out for supplementary contributions. In the meantime, however, there are other tasks on which the foundation could busy itself. The recent study on the mobility of academic scientists within the European Community (see *Nature* 23 October) made much of the lack of movement between European countries (as distinct from the now common journeys across the Atlantic). As things are, several of the foundation's committees organize research workshops, symposia and the like in their narrow fields, but the scale of such activities within Europe is nothing like enough, in spite of the efforts of the European Molecular Biology Organization and the federations of specialist scientific societies that have sprung up. More ambitiously, there is a case for asking the foundation to share in the administration of the European exchange fellowship scheme, so far in the hands of national academies but much in need of escaping from the strictly bilateral basis on which it was set up eight years ago. (With a little luck, an international organization might even be able to tap supranational sources of funds, the European Community for example.) Finally, there are some crucial housekeeping tasks not at present being done (or done well) for European science. There is, for example, a crying need for a directory showing who does what in European research, some means of knowing who is where and a systematic way of telling which national organizations spend how much. Coherently compiled statistics of the working (and output) of European university systems would be a great boon. If one immediate objective is to ensure that the smile remains visible and the rest of the animal inconspicuous, this must surely be the best way forward.

Backlash against DNA ventures?

More thoughts at Harvard on business link

Washington

Harvard University's announcement that it is considering setting up a company to exploit the results of its recombinant DNA research (*Nature* 30 October, p.769) has provoked a backlash on university campuses that may lead Harvard to drop the idea. The company is seen by some as a threat to academic freedom.

University administrators have, in general, welcomed the Harvard proposal, particularly since many of them are already discussing similar ideas of their own but have not yet made them public.

Yet at the same universities both scientific and non-scientific faculty members have warned of the dangers of steering basic research towards commercial objectives, suggesting that there are elements in the Harvard proposal which make it qualitatively different from more conventional arrangements for linking university research and its industrial applications.

One group which has taken up the issue is the academic freedom committee of the American Civil Liberties Union (ACLU). The committee has set up a subcommittee to draft a possible policy statement for ACLU to consider on guidelines to cover relationships between universities and industry.

"I do not think that you can divorce the universities from industry, but the contribution that we can make is to set guidelines down for the future, for which we feel both faculty and administrators would be appreciative, as they have been in the past on issues such as military research", Dr Bernard Belush, chairman of the academic freedom committee, said last week.

Other universities considering new ways to profit from their research are going through similar debates to that at Harvard, where more than half the faculty members were reported by the *Wall Street Journal* to have told Harvard president Derek Bok that they were opposed to the idea.

At Yale University in New Haven, for example, the faculty has asked for a revision of proposals which have been put to it for encouraging greater links with industry, and has expanded the committee looking at these proposals — which might, as at Harvard, include the idea of setting up a research development company — to include more non-scientists, who had been among the critics of the original proposal.

A similar debate is expected at the

University of Michigan at Ann Arbor, where there is a proposal to link up with a local pharmaceutical company, Warner Lambert, to provide training facilities for the company's research personnel.

Dr Charles Overberger, vice-president for research at Ann Arbor, said last week that although there were no firm plans to set up a separate company, talks had been held with a venture capitalist, Mr Herbert Doan, an ex-chairman of the Dow Chemical Company.

At Stanford University in California, where much of the original recombinant DNA research was carried out, the debate is already at an advanced stage. A proposal similar to Harvard's was made several years ago that the university should become more directly involved as an investor in exploiting its research results, rather than merely offering patent rights.

The proposal was, however, shelved after strong opposition from several faculty members, including Professor Paul Berg of the university medical school's biochemistry department.

Rather than getting directly involved, the department is now setting up an industrial affiliates programme under which companies which sign up will be able to send representatives to one seminar a year on the department's work, to receive one visit a year from a faculty member to discuss the company's research and to send a representative to Stanford to discuss this research with faculty members.

In return, the company will give \$12,000

annually to the university, which the department hopes will be used to support junior research posts. Fifteen companies — including Cetus, Genentech and several major pharmaceutical firms — have already signed up, and the department has a growing waiting list.

Dr Donald Kennedy, who took over as Stanford's president earlier this year after spending four years as head of the Food and Drug Administration in Washington, admits that the new commercial pressures on biological research have thrown up a variety of problems which biomedical researchers have never previously had to face. "Quite possibly it is time for some leadership to be sought or to well up in universities on this point as it has done on other points in the discussion of the social responsibilities of scientists", says Dr Kennedy. "Perhaps it is time for another Asilomar, but it ought to come from the people who are doing the science."

At Harvard, university officials insist that they have gone to considerable lengths to prevent any potential conflict of interests between the academic goals of the institution and its investment policies. For example, no university administrators would be directly involved in investment decisions taken by the new company, which would be completely separate from the university. Although the university would be a minority shareholder, it would not be investing any of its endowment, since the shares would be provided in return for access to research results.

Science Research Council comes clean

The UK Science Research Council (SRC), which could not say (see *Nature* 13 October) how many research grants were outstanding at the end of the last financial year because of a computer oversight, has now produced reliable figures. According to the revised account, 3,933 grants were current on 31 March 1980, of a total value of £120.5 million. This is an increase in number of 15 per cent and in value of 34 per cent over the previous year.

The increase is still substantial and unplanned. Indeed, it is the first major increase of grant allocations for many years. It is only partly accounted for by the "equipment round", a £7.5 million handout for major items of equipment in universities which took place in the summer of 1979.

Last week the Department of Education and Science (DES) came to SRC's aid, allowing the council to transfer £5.7 million of its savings on foreign subscriptions, due to the strong pound, to other accounts. SRC can thus pay for some of this year's increased expenditure on grants. Even so, the council was due this week to seek another £2 million of savings on current spending, on top of £4 million

worth of cuts already made.

SRC will therefore scrape through the current financial year without requiring extra finance from DES — but the year to follow (April 1981 to March 1982) looks increasingly bleak. The pound seems unlikely to rise further against the European currencies in which foreign subscriptions are mostly paid, and indeed it may fall if interest rates in London money markets are reduced. Thus the £5.7 million found on exchange rates this year may vanish or even become negative next year, while the grant commitment will remain obstinately high.

Next year, accordingly, the council will have to seek an increase in its funding above the £175.3 million included in the provisional 1980 five-year forward look, to pay for its support for the universities last year. There is no reason to suppose that the other dependants of the Advisory Board for the Research Councils will be sympathetic. The council's finances are, however, unlikely to be seriously affected if it is required to refund the price of its annual report (HMSO, £7.50) to purchasers complaining that the figures it contains are wrong.

Robert Walgate

Whether this will be sufficient to answer the critics remains uncertain. While scientists were voicing their fears about academic freedom, several investment houses were saying in public last week that they thought it was a bad thing for a university to get involved in commercial operations. Dr Bok's announcement about whether the university intends to go ahead with plans to set up the new company is therefore awaited with great interest.

David Dickson

European science

Foundation stones

Strasbourg

The European Science Foundation ended its annual general assembly here last week (13 November) in a vaguely hesitant frame of mind. Unlike previous assemblies, this was less a general amplification of the enthusiasm for international collaboration of member organizations (now 47 research councils and learned academies from 18 countries) than a reluctant recognition that the next six years may be more difficult than the past — and first — six.

Part of the hesitancy stems from changes of personnel. Lord Flowers, rector of Imperial College, London, and president of the foundation for the past six years, handed over at this year's assembly to Professor H. Curien, director of the CNRS (Centre National d'Etudes Spatiales) in Paris, while Dr. John Goormaghtigh succeeded Dr Friedrich Schneider as secretary-general only a year ago. But the member organizations seem also to be preoccupied with domestic funding problems and thus less able to think expansively than in recent years.

Several of the foundation's recent initiatives appear also to be at critical, even

anxious, stages. Thus the development of a plan for what is called the European Synchrotron Radiation Facility, on which a foundation study group has been working for the past three years, is in the air. At the general assembly in 1979, the retiring secretary-general, Dr Schneider, was asked to spend a year sounding out European governments on their willingness to contribute towards the cost of the machine, estimated at something of the order of \$100 million.

Dr Schneider appears to have found this a disappointing year. Potential contributors to the cost of the machine have taken the view that haste might jeopardize eventual construction. Some governments also appear to be suspicious that the foundation is usurping their role of making treaties among themselves for multinational projects. The study group (under Professor Y. Farge of Orsay) will, nevertheless, remain in being during 1981, concentrating on the further refinement of its proposals.

The foundation is also despondent about the prospects for space research in Europe. A year ago, one of the foundation's published reports took the European Space Agency to task for the degree to which its then programme was biased to the development of the Ariane launching system, to the detriment of plans for launching scientific satellites. At this assembly, delegates were told that there has been no substantial improvement. Plans for launching European scientific satellites are so sparse that scientists who are members of space research groups in Europe may have to wait about five to seven years between successive experiments. Professor Curien told the assembly that this might be inevitable in studies of "the reproduction of the elephant, or of silviculture", but that it was

an entirely artificial circumstance in space research.

The foundation is not, however, entirely frustrated. Its programme on polymer structure is going well (see box), while the assembly agreed to launch a number of new projects in the social sciences while continuing most of its past successes, the programme of training in "brain and behavioural research", for example.

On more general questions, the foundation has decided to keep in being the Liaison Committee on Recombinant DNA Research even though, according to a statement from the committee, the need for attempts to coordinate containment guidelines has melted away. The focus of anxiety has moved to what are called "second generation" issues, including the pace and the manner in which recombinant DNA techniques are being commercialized with the possible consequential risks to academic research. Dr Philip Handler, president of the US National Academy of Sciences, said in his invited address to the assembly that academics forming commercial links with DNA companies were "creating a great deal of difficulty for the others working in this field".

The foundation has also issued a statement asking that national governments planning new legislation on the confidentiality of computerized data banks should also bear in mind the value of such data, suitably shorn of identification, for the research community.

The assembly approved the foundation's budget for 1981, which exceeds FF5 million for the first time, and which does not include the cost of the special research projects (called "additional activities") to which member organizations contribute on a voluntary basis.

Lord Flowers estimated, in his final address to the assembly, that the total expenditure on the foundation's activities amounted to less than 0.1 per cent of the budgets of member organizations, and hoped that the next six years would see a tenfold increase in this proportion in the next six years, as well as a further extension of the foundation's sphere of interest. The academy of Finland was admitted as a full member of the foundation with effect from 1981.

Fusion research

Livermore looks up

Washington

After a period of considerable uncertainty, scientists in the two main fusion energy programmes at the US Department of Energy's Lawrence Livermore Laboratory in California are confident that recent technical advances are given a significant boost to the chances of being able to exploit nuclear fusion as a commercial source of power.

In the magnetic confinement programme, successful tests with so-called tandem mirrors, which could be used to

Something ventured, something done

Both the new research projects approved by the 1980 assembly are in the social sciences. One, under the rubric of "comparative law", and to which ten countries have agreed to contribute, includes an attempt to define and compare the medical responsibility within Europe. The study will include the responsibilities of physicians and drug firms towards patients and the influence of medical insurance on the behaviour of physicians. There is also to be a European study of procedures for summary jurisdiction in civil and other courts.

The social scientists are preoccupied with migration and language, and have won approval for two other research projects — the problems of language acquisition by adult migrants within Europe (of whom there are estimated to be 11 million) and a series of research workshops on the human and cultural aspects of migration within Europe.

There is also to be a one-year study of

problems of technical innovation and social change, chiefly so as to identify fields in which coordination by the foundation might be beneficial.

Continuing projects that appear to be flourishing, or nearing fruition, include:

- *Chinese studies.* A handbook of Chinese literature in the period 1900-49 is nearing completion, while a descriptive catalogue of the body of fifteenth century Chinese literature known as the *Tao-Tsang* is well under way.

- *Brain and behaviour research.* The foundation plans to spend in 1981 a total of FF1.3 million on this programme of training awards and travel grants.

- *Taxonomy.* The *ad hoc* group on European taxonomy is to be disbanded, and its final report published in March next year. The draft report claims to have made substantial progress towards understanding the difficulties of coordinating European taxonomic nomenclature.

"plug" the ends of a solenoid containing plasma, have led to the decision to include the mirror devices in the Mirror Fusion Test Facility (MFTF) at present under construction. Work is already under way on plans for a new, larger machine, tentatively referred to as the Tandem Mirror Next Step (TMNS), which could become a serious rival to the tokamak design.

The programme based on inertial confinement techniques is also looking well. Recent experimental data indicate that solid state lasers made from vanadium-doped magnesium fluoride crystals have characteristics, high energy efficiency chief among them, which may overcome the difficulties experienced with the use of more conventional solid state lasers to ignite a deuterium-tritium pellet.

The tandem mirror is a device first proposed in 1976, when experimenters were experiencing considerable difficulty in preventing plasma from escaping its containing magnets along field lines which are open rather than closed (as they are in a tokamak). Its ancient ancestor is the machine called DCX, developed at the Oak Ridge National Laboratory in the 1950s.

Proposed simultaneously by Ken Fowler and Grant Morgan at the Livermore Laboratory, and Soviet scientists at Novosibirsk, the mirror acts by using two magnets to contain a plasma in an electrostatic potential well, the plasma being positively charged because electrons are able to escape faster than protons.

Small-scale experiments were carried out by research workers at the University of Tsukuba in Japan last year, demonstrating that, in principle, a tandem mirror designed to contain a plasma in this way does create the predicted potential well.

Subsequently, Livermore scientists have concluded a series of experiments on the tandem magnet experiment (TMX) in which they have been able to demonstrate that a solenoid with such plugs at each end can contain a heated plasma at a mean energy of 0.2 keV per atom, as compared with 13 keV within the end mirrors.

The attraction of a commercial fusion reactor based on this design would not only be the easier design tasks presented by a straight solenoid rather than a toroidal tokamak, but also that, once the end plugs have been produced to the tandem mirror design, the length of the solenoid between them can be adjusted at will to provide the power characteristics required.

The US Congress has already been sufficiently impressed to approve substantially increased funding for the MFTF now being built.

Its success or otherwise will determine whether the department should proceed to the next logical step, the TMNS, Livermore scientists feel that magnetic mirror designs could catch up with tokamak technology, the most likely design for a Fusion Energy Device (FED), and which the department hopes to build over five to ten years.

Meanwhile, those working on the

inertial confinement techniques are hoping that the promise of vanadium-doped magnesium fluoride lasers may overcome doubts about the ultimate potential of solid state lasers as fusion drivers.

Most of the research so far has been done on glass lasers, initially with red light and more recently, following the development of techniques for doubling the wavelength, with green and blue light which has a greater ability to focus energy on a small deuterium-tritium pellet.

Preliminary evidence suggests that, in contrast with the conventional krypton fluoride laser, which only has an efficiency of between 5 and 7 per cent, a V:MgF₂ solid state laser would have an efficiency of between 5 and 10 per cent. With other advantages, this would help to reduce the cost of electricity from an estimated \$300 million to \$100 million per megajoule.

Livermore scientists are now carrying out further experiments to determine whether crystals with such characteristics can in fact be grown for use in lasers. If so, solid state lasers will remain a serious contender in the fusion stakes.

David Dickson

Radioactive waste

Brussels helps

The parallel research programmes of the European Economic Community (EEC) and Canada on the storage of radioactive waste are to be linked. A five-year agreement was signed in Brussels on 3 November between the European Atomic Energy Community (Euratom) and Atomic Energy of Canada Ltd. It follows on from the original Euratom/Canada agreement of 1959.

The new agreement relates particularly to the evaluation of the environmental impact of the storage of wastes in hard rocks, and to collecting data on the development of safe waste storage systems. There has already been considerable cooperation between scientists in these fields, so the agreement puts this on a more formal basis and opens the way for even closer cooperation.

Initially the agreement will lead to exchanges of technical information, organization of joint scientific meetings, and exchange visits of scientists between Canadian and European laboratories.

The agreement is the first in this field between the community and an outside country. It is likely to be a pointer to a trend towards more international cooperation, and the United States is already showing interest in closer cooperation with the EEC.

The community's own research programme was unveiled with a great fanfare in Luxembourg in May this year. The community's contribution, US\$130 million, is planned to extend over four years. Of the total, \$100 million will be contributed on a fifty-fifty basis to national research programmes, while \$30 million will be spent at the joint research

centre at Ispra.

Research on the disposal of high-level wastes is well advanced in France, West Germany and the United Kingdom. In each case, the objective is to embody high-level radioactive wastes in some solid form — France and the United Kingdom are chiefly interested in vitreous solids, while West German studies have also taken account of bitumen-like materials.

In several member states, investigations are also under way to identify disposal sites for solidified radioactive waste. France and the United Kingdom are exploring granite formations, West Germany salt formations while Belgium and Italy are concentrating on siliceous (clay) formations.

Jasper Becker

Waste disposal

UK goes French

British Nuclear Fuels Ltd (BNFL) whispered last week its intention to use the French AVM process for vitrifying highly active nuclear waste emerging from the Windscale reprocessing plant, thus bypassing the British "Harvest" vitrification research programme which has been under way at Harwell. There was no official announcement by BNFL: rather, the news emerged during a routine meeting of the Windscale local liaison committee, a group convened to keep the local community aware of developments at the plant. The leak was intentional, and confirmed rumours that BNFL had chosen AVM some time before.

A full commercial agreement between BNFL and Cogema, the French company which developed AVM, has not yet been reached, and if Cogema's price is too high the deal may still fall through. As presently envisaged, it would entail constructing an AVM (*atelier de vitrification a Marcoule*) plant at Windscale in the late 1980s, under licence from Cogema.

BNFL will not say at this stage why it has chosen AVM, though the reasons are not far to seek. Both the French and the British began work on vitrification in the late 1950s but work halted in Britain in 1966 when the principle had been demonstrated. (It was felt at the time that it would be possible to store liquid high active waste indefinitely in tanks.) Political pressure led to the setting up of the Harvest programme (Harwell vitrification engineering study) in 1973, but by then Cogema had established a seven-year technological lead.

While the Harvest team at Harwell is even now using only simulated nuclear waste, a plant called PIVER was built at Marcoule eleven years ago to handle the real stuff in commercial quantities. PIVER ran from 1969 to 1973, and produced 12 tonnes of vitrified waste containing 5 million curies of activity — roughly speaking the arisings of a 1 GW nuclear power plant working over the same period.

PIVER was a batch process, like



Product of Harvest

Harwell's Harvest; its successor AVM, incorporated in a plant at Marcoule which began operation two years ago, is continuous, allowing — it is claimed — a roughly twofold saving in capital equipment costs for the same waste throughput. The AVM plant is remotely controlled, and is constructed from units none larger than 500 kg for ease of handling and replacement. The French process is thus many years ahead of its British — or any — rival.

Nevertheless there is still a question mark over BNFL's haste to conclude an agreement with Cogema. The decision to abandon Harvest is as much a political as a commercial one. At present, about 1,000 m³ of high active waste is stored in liquid form in tanks at Windscale, and BNFL still considers that it could remain in these tanks, or new ones, for some decades. Further tanks could be built for the oncoming nuclear power programme, but it is felt that public opposition to the programme — which will be expressed in the promised PWR enquiry — may be less if vitrification work is seen to be progressing apace. The argument, however, turns a

The differences

High active waste emerges as a liquid from the reprocessing of spent fuel: it is a solution in concentrated nitric acid of the nitrates of actinides and fission products. The object of vitrification is to immobilize the radioactive elements. In the Harvest process, which is only at the pilot stage, a simulated waste is mixed with the components of a borosilicate glass, poured into a tube, evaporated and glassified in one step in a furnace. The tube becomes the primary container for the glass. The process is a batch process. In the AVM technique, a high active liquid stream is evaporated and calcined (converted to oxides) in a rotary calciner. Each hour, the calciner accepts 40 litres of liquid waste and converts it to brown granules, which leave the furnace continuously. The granules are mixed with borosilicates and melted in a stainless steel container. Every eight hours, the container pours out 140 kg of glassified waste into storage drums. The drums are moved out, lids are welded on, and the drums moved to storage caves — all automatically.

blind eye to local opposition to drilling tests, to find sites for future long-term disposal of the glass.

At Harwell, Dr Ron Flowers, who heads the chemical technology division that developed Harvest, says that he expects to close work on the project by next March, once experiments are completed and final publications prepared. Loss of the Harvest contract — which was paid for by BNFL, the Central Electricity Generating Board, and the Department of the Environment — will not be a serious blow to the division, he says.

Robert Walgate

Soviet dissidents Another taken

Viktor Brailovskii, the Moscow Jewish mathematician who regularly hosts the Sunday seminars for refusenik scientists, was arrested last week, reportedly under Article 109/1 of the Criminal Code which deals with slander of the Soviet Union and the socialist system and the circulation of "publications known to be false".

News of the arrest, coming shortly after the opening of the Madrid review conference on implementation of the Helsinki Final Act, has caused consternation among his colleagues and friends abroad, who see it not only as a threat to Brailovskii himself, but also an adverse omen for Soviet intentions regarding "Basket 3" of the Final Act, which deals with human rights.

Brailovskii, who is a specialist in cybernetics and computer programming, first applied to emigrate to Israel in 1972. His application was refused, and he was then dismissed from his academic post. Since then, he has had no chance of obtaining any job in science.

It was to help such "refusenik" scientists keep up their academic interests that the Sunday seminars were founded in 1973 by Aleksandr Voronel'. When Voronel' and then his successor as seminar leader Mark Azbel' finally managed to emigrate, Brailovskii and his wife Irina (also a mathematician) became hosts to the seminar.

A recent Western visitor reports that in spite of a total lack of access to scientific facilities, including journals, the standard of discussion remains high, and there is considerable participation by young Jewish scientists who are attempting to carry out some kind of postgraduate study on their own.

A frequent visitor to the seminar, the traveller said, is Evgenii Chudnovskii of Khar'kov — a city where the refusal rate has always been particularly high — who is now trying to organize what he calls a "Jewish university" for young refuseniks.

According to Mrs Brailovskaya, the seminars will continue in spite of her husband's arrest. The citation of Article 109/1 presumably refers to the *samizdat* journal *Jews in the USSR*, of which

Brailovskii was for a time editor. However, this journal ceased publication in summer 1979, and the long delay before an accusation on this charge seems unreasonable.

German science

New man arrives

Germany has a new Minister for Science and Technology: 43-year-old Andreas von Bülow, who since 1976 has been Parliamentary Undersecretary of State for Defence. The Bundesministerium für Forschung und Technologie BMFT is Dr von Bülow's first ministerial appointment — as it was for his predecessor, Dr Volker Hauff. Dr Hauff now moves on to the Ministry of Transport, which is considered to be a more senior political appointment.

Dr von Bülow has had no time yet to form his policies for the BMFT, but he is expected to lean more to the right of his party, the SPD, than Dr Hauff, who attempted to use the ministry as an



instrument of economic policy (through its involvement in technology). But Dr Hauff — after seven years in some capacity at the BMFT — was said to be becoming impatient at the inertia of the research community, and particularly the big science institutions like the Karlsruhe nuclear research centre and others which had grouped themselves into a powerful lobby, the Arbeitsgemeinschaft der Grossforschungseinrichtungen (AGF). It remains to be seen what Dr von Bülow will do with that. His ministry plays a coordinating role in German science, akin to that of the Delegation Générale à Recherche Scientifique et Technique in France; but it controls a much larger budget — around DM 6 billion (£1.25 billion). This is the bulk of the federal finance applied to R&D Germany, another DM 3 billion or so being under the direct control of other ministries. (A further DM 6 billion is spent by the state governments, mostly to support universities, and DM 16 billion by industry.) The BMFT, along with the state governments, supports the institutions of the AGF, the Max Planck Gesellschaft (which runs the Max Planck institutes), the Deutsches

Forschungs-gemeinschaft (which funds research projects in the universities), the Fraunhofer Gesellschaft (the equivalent of Max Planck, but for applied science), and other science promotion bodies. The ministry also promotes research and innovation in industry through a number of mechanisms.

Robert Walgate

Indian environment

Shandi inverted

Bangalore

Mrs Gandhi's new government seems to have taken the environment to heart, to judge from the spate of decisions since the election at the beginning of the year and the appearance of the environment in the newly published Sixth Development Plan.

In recent years, environmental abuse in India has increased alarmingly. Urbanization, industrialization and river pollution have all conspired to make a mockery of the United Nation's slogan of "development without destruction" in a country renowned for its religious regard for the variety and greatness of nature. The repercussions of this indiscriminate assault on the ecological system are telling:



Barren hill

landslides and flash floods in the Himalayan foothills have caused heavy loss of human lives and property, top-soil erosion and growing aridity in the Ganges plains have had severe consequences for agriculture.

Successive Indian governments during the past 30 years have shown little awareness of the cardinal importance of environmental planning in developmental activities. It was only after Mrs Gandhi's return to power in early 1980 that the environment assumed significance.

Since then, work on the controversial Silent Valley hydroelectric project in the tropical rain forests of southern India has been suspended and the rate of deforestation in the Himalayan foothills has been

greatly reduced.

It is a measure of the growing public ecological awareness that the last session of the Indian parliament witnessed perhaps one of the liveliest debates on the environment. Intervening in the debate, Mrs Gandhi made it clear that the nation must not repeat its earlier mistake of allowing industrial projects, despite their economic importance, to damage the delicate environmental web.

During the debate, a ruling party member, Dr Karan Singh, expressed shock over the "ruthless denudation of Himalayan vegetation by corrupt politicians, corrupt officials and corrupt contractors". He further lamented that even the Ganges had now been contaminated.

Another member, Mr Digvijaya Narain Singh, said that, as a result of top-soil erosion, 90 million hectares — equal to 28 per cent of the total land area — had now become practically barren. He urged that there should be a separate department concerned with land use, forestry, wild life, pollution control, marine ecosystems and the promotion of environmental protection.

This suggestion seems likely to be incorporated into the government's long-term environmental strategy, which will be based on the recommendations of a 14-member committee set up to suggest legislative measures and administrative machinery for environmental protection. The committee, headed by Mr Narayan Dutt Tiwari, deputy chairman of the Planning Commission, has come out strongly in favour of immediate coordinated action at both central and state levels to give environmental protection a crucial place in the country's programmes and policies.

The Indian government is also contemplating introducing a bill on the prevention of air pollution in which would be based on proposals put forward by an expert committee appointed in 1978 to study air and water pollution in the urban areas of India. Air pollution has become a major health hazard in cities such as Bombay, where the content of carbon dioxide in the atmosphere has been increasing by 4.2 per cent a year due to the 65 tonnes of dust ejected into the air each day by industry and vehicles.

B. Radhakrishna Rao

Nuclear fallout

Weapons are worst

States with substantial nuclear industries should do everything they can to avoid strategic nuclear attacks. That is one of the inevitably ironical conclusions of a study of catastrophic nuclear radiation releases by Steve Fetter and Kosta Tsipis of the Program in Science and Technology for International Security at Massachusetts Institute of Technology.

The chief objective of the study has been to compare the long-term consequences of nuclear weapons exploded in the air and on the ground, catastrophic releases of radioactivity in reactor accidents and the explosion of nuclear weapons on or near nuclear reactors or reprocessing plants. Although it is recognized that the immediate effects of weapons explosions as distinct from reactor accidents will consist largely of the death of people and the destruction of property, the calculations are aimed at estimating the area of surrounding land that will be unfit for human habitation after an explosion or some other catastrophe.

For the purposes of the calculation, the authors assume that half the energy released in a one-megaton thermonuclear explosion is provided by fusion reactions and the remainder by the fission induced in a surrounding blanket of uranium-238. For weapons burst in the atmosphere, radioactivity from the fission of uranium-238 will be the sole source of long-term contamination, and much of its debris will be distributed by stratospheric processes.

The areas rendered uninhabitable after a nuclear explosion are sensitively dependent on the criteria used for deciding what doses of radioactivity are acceptable. For a ground burst weapon whose debris is spread in the lower atmosphere by a 15-mile per hour wind, the authors calculate that 5,700 square miles of land would be uninhabitable for a year if doses greater than 2 rem per year were considered unacceptable, but that fewer than 50 square miles would be uninhabitable a year after the explosion if 100 rem per year were taken as the cut-off dose.

Inevitably the consequences of the explosion of a nuclear weapon on or near a reactor are more startling. Using the somewhat stringent criterion of a limiting dose of 2 rem per year, the authors conclude that 50,000 square miles of land would be uninhabitable for a year. The destruction of a waste storage facility would have still more horrendous consequences, putting 64,000 square miles out of action for a year. By comparison, the report concludes, the consequences of a reactor melt-down and a subsequent release of radioactivity would be comparatively small, sterilizing only 900 square miles for a period of one year.

The obvious but impractical import of these calculations, Fetter and Tsipis say, is that countries seeking to avoid devastating damage from distributed radioactivity in a nuclear war should either dispense with a nuclear industry or build reactors and reprocessing plants underground. They also point out that because the permanent damage done by a single nuclear weapon detonated on the ground is so much greater than that in the "worst conceivable nuclear reactor accident", it is hard to understand the greater anxiety of the general public about the risks of civil nuclear accidents than with the consequences of nuclear war.

CORRESPONDENCE

Museum of errors

SIR — Two years ago (*Nature* 275, 683; 1978), I questioned the wisdom of what is happening at the Natural History Museum in South Kensington. If a national museum is concerned with aspects of social engineering, by promoting concepts that happen to be current in the present climate of opinion, are there not sinister implications? I was especially alarmed by the museum's new exhibition scheme, and asked that sufficient pressure should be brought to bear to "curb the activities of the Public Services Department and to ensure the survival of the museum's reputation for scholarship in its public galleries".

Since then, time has passed. It is no longer a question of raising the alarm but simply of reporting what has already happened.

Two areas of the museum's work have already succumbed: dinosaurs in 1979 and fossil man in 1980. Both the new exhibits are simply vehicles for the promotion of a system of working out relationships known as cladistics. The accompanying booklets *Dinosaurs and their living relatives* (1979) and *Man's place in evolution* (1980) explain with startling clarity the essence of cladistics. In both books the fundamental assumptions are spelt out unequivocally. "First we assume that new species arise when one species splits into two. This assumption allows us to test the relationship we suggest, because it means that every species must have a closest relative. Second we assume that none of the species we are considering is the ancestor of any of the others." It is axiomatic, therefore, that no species in the fossil record can be considered ancestral to any other nor can one species evolve directly into another.

With regard to both dinosaurs and fossil man, it is evident that the application of cladistics is quite inappropriate. The well attested sequence of human fossils representing samples of succeeding populations has, until the Natural History Museum's latest exercise, been taken as a classic example of the gradual evolution of a single gene pool. Certainly there is not any serious doubt about *Homo erectus* being directly ancestral to *Homo sapiens*.

Yet the concept of one species being directly ancestral to another is contrary to the rules of cladistics. So we read in the section on *Homo erectus* (under the heading "Not our direct ancestors"), that "The *Homo erectus* people were not quite like us. . . the *Homo erectus* skull has several characteristics that the modern skull does not share. Because of these special characteristics, we think that the *Homo erectus* people were not our direct ancestors".

But then on the opposite page is a photograph of the Petralona skull, from Greece, which the author considers an example of *Homo sapiens* or *Homo erectus* because of its mixture of features. This particular skull makes nonsense of the entire methodology being promoted in the books and exhibition. According to the stated assumptions of cladistics none of the fossil species can be ancestral by definition. This presents the public for the first time with the notion that there are no actual fossils directly antecedent to man. What the creationists have insisted on

for years is now being openly advertised by the Natural History Museum.

The scientists on the museum staff, be they experts on dinosaurs or on fossil man, have had their scientific judgement over-ridden by the Department of Public Services. What exactly is the cladistic framework to which the Public Services Department is so fervently dedicated? Why is there such a fanatical insistence that data should be presented within such a framework?

And why should there be a deliberate policy that involves the removal from the public gaze of important and scholarly exhibits in the museum such as the Insect Gallery and the Fossil Mammal Gallery? Is it because they provide too dramatic a contrast with the propaganda of the new-style exhibits?

The questions that should arise in everyone's mind are: what is this all about, what actually is going on and what is behind it all? The answers can be found by reading the literature of cladistics. The tenor of this is seen in its abuse of E. Mayr and G.G. Simpson, and indeed of Charles Darwin himself, because of their firm adherence to the concept of gradualism and to the idea that the processes that can be observed at the present day, when extrapolated into the past, are sufficient to explain changes observed in the fossil record. The synthesis of population genetics and palaeontology presented by Simpson in his two seminal works *Tempo and mode in evolution* (1944) and *The major features of evolution* (1953) is anathema to cladists.

The next question is why should the notion of gradualism arouse passions of such intensity? The answer to this is to be found in the political arena. There are basically two contrasting views with regard to human society and the process of change through time: one is the gradualist, reformist and the other is the revolutionary approach. The key tenet of dialectical materialism, the world outlook of the Marxist-Leninist party according to J.V. Stalin, is in the recognition of "a development in which the qualitative changes occur not gradually but rapidly and abruptly, taking the form of a leap from one state to another" (Engels). This is the recipe for revolution. If this is the observed rule in the history of life, when translated into human history and political action it would serve as the scientific justification for accentuating the inherent contradictions in society, so that the situation can be hurried towards its appropriate "nodal point" and a qualitative leap supervenes.

With regard to evolution and the fossil record, neither Engels nor Lenin, both of whom discussed the subject at length — to their great credit — insisted upon a pattern of such qualitative leaps, they were merely content to see in evolution and the fossil record evidence of change, albeit gradual.

This has always been a matter of some disquiet for Marxist theorists. If it could be established that the pattern of evolution was a saltatory one after all, then at long last the Marxists would indeed be able to claim that the theoretical basis of their approach was supported by scientific evidence. Just as there are "scientific" creationists seeking to falsify the concept of gradual change through time in favour of catastrophism, so too there are the Marxists who for different motives are equally

concerned to discredit gradualism.

What is going on at the Natural History Museum needs to be seen in this overall context. If the cladistic approach becomes established as the received wisdom, then a fundamentally Marxist view of the history of life will have been incorporated into a key element of the educational system of this country. Marxism will be able to call upon the scientific laws of history in its support, with a confidence that it has previously enjoyed.

This is the course of action to which the authorities of the Natural History Museum seem to have committed themselves either unwittingly or willingly.

L.B. HALSTEAD

Department of Zoology and Geology,
University of Reading, UK.

Refrigerator brew

SIR — One of the problems associated with Legionnaires' disease lies in the tracking down of the source of infection. Air conditioning systems seem to be prime suspects. So far as I am aware, nobody has thought of looking behind refrigerators. Having had occasion to do this recently, I was taken aback to discover that a self-defrosting refrigerator may provide the ideal breeding ground for a wide range of microorganisms.

At the back of a self-defrosting refrigerator there is a reservoir to collect the water as it is discharged. This water remains stagnant, gradually evaporating until it is replenished by the next automatic defrosting. With some designs the cooling coil passes through this reservoir to hasten evaporation. The result is a warm broth ideally suited to the proliferation of microorganisms. After a year or two stood around in a hospital environment such reservoirs might well harbour some very interesting cultures indeed. Even the ordinary household must pick up some pretty virulent infectious agents from time to time.

I have now taken the precaution of adding a dash of disinfectant to the brew. Hospital authorities might care to adopt the same simple measure.

ROBERT T. GREEN

15 Clifton Gardens,
London NW11, UK

Frequency tripling

SIR — Your correspondent's comments on the US inertial-confinement fusion (ICF) programme contains a minor inaccuracy (*Nature* 16 October p.573). University of Rochester scientists have, in fact, frequency-tripled the output of their glass laser thus converting the 1.054 μm light to 0.3513 μm radiation at high efficiency. Experiments using frequency-doubled light, at 0.53 μm have been done or are underway at many laboratories, notably the Science Research Council Rutherford Laboratory, the Centre d'Etudes de Limeil (France), GRECO-Interaction Laser Matière, Ecole Polytechnique (Palaiseau, France), KMS Fusion Inc., and the Lawrence Livermore National Laboratory. The Ecole Polytechnique group recently has reported on experiments using 0.26 μm radiation, the 4th harmonic of 1.06 μm light.

La Jolla Institute,
La Jolla, California

PETER HAMMERLING

NEWS AND VIEWS

What prospects for an X-ray laser?

WHAT are the chances that one day there will be machines capable of producing short but intense pulses of virtually monochromatic X rays, for all the world like lasers that function at the short-wavelength end of the spectrum? And if such machines can be built, what purposes will they serve? The notion that a machine with such a specification might be the basis of yet another military weapon will no doubt quickly capture the imagination of newspaper-headline writers. In reality, of course, the Earth's atmosphere is as opaque, more or less gram for gram, as are other materials. For those who brood about "death rays" for the benefit of science fiction literature, what matters (or should matter) is not the wavelength of the radiation but the amount of energy that can be delivered by a radiation beam and the efficiency with which it is absorbed by the target. Reflection will also show that the present attempts to develop X-ray lasers, centred as they are on machines that differ very little from high-energy particle accelerators, suggest that X-ray lasers — if they can be built — will hardly be the kinds of portable devices of which generals and their staffs are fond. In other roles, however, and especially in strictly scientific applications, devices of this kind, however cumbersome, would be of great benefit. So what are the chances that they will eventually be constructed?

This question has obviously perplexed those within the European Science Foundation who have been concerned with the project to plan a European facility for synchrotron radiation. At the general assembly of the foundation last week in Strasbourg, those responsible for this project, which is essentially a scheme for building an electron storage ring operating at 3 GeV and designed to produce copious amounts of synchrotron radiation, let it be known that they are undeterred by the chance that an X-ray laser may eventually make their own project unnecessary. In this phlegmatic view, the planning group is surely on safe ground. The chances that it will be possible to construct machines operating as lasers in the X-ray range within the lifetime of the proposed European synchrotron machine are small. Further ahead, however, there is a reasonable chance that something may be possible.

The starting point for this speculation is the wave of excitement generated in the past few years by the concept of the "free-electron laser", an innovation largely due to Dr J. Madey at Stanford University in 1976 (Elias, L.R. *et al. Phys. Rev. Lett.* **36**, 717; 1976). The working principle of these devices is a curious amalgam of old-fashioned 1940s technology and the more recent sophistication of the particle-beam technologists (see J.N. Elgin *Nature News and Views*, 1 May 1980, and references therein). Suppose a beam of electrons travelling along a tube is subjected to a sequence of transverse magnetic fields, alternately in one direction and the opposite. At each step, the electrons (accelerated as they will be along the third cartesian axis) will radiate electromagnetic energy. As Maxwell's equations have it, a beam of electrons processed in this way should be a source of a radiation beam directed along the axis of the system. Moreover, the wavelength of the radiation should be determined to a first approximation only by the energy of the electrons and the geometrical periodicity of the magnetic field. The more frequent the alternation of the direction of the magnetic field along the axis of the beam, the shorter the wavelength of the radiation, which is (or should be) related inversely to the square of the energy of the electrons. The planning of the proposed European synchrotron radiation facility is based almost entirely on the notion that a circulating beam of stored electrons should be subjected to a sequence of such corrugated magnetic fields, called "wigglers" in the trade.

The novel development at Stanford University was the recognition that a beam of electrons travelling through a corrugated magnetic field should also behave as if it were the working material of a laser. In other words, if an electron beam traversing a wiggling magnetic field is also exposed to radiation of the frequency characteristic of the geometry and the electron energy, radiation at that same frequency will be stimulated and, if there is a pair of mirrors somewhere, energy from the mechanical motion of the electrons will be converted into a beam of radiation of wavelength determined in advance only by geometry and electronic energy (with the actual strength of the magnetic field as a second approximation correction). So, would-be designers of X-ray lasers will say, why not go the whole hog, pushing the electron energy as high as possible but reducing the geometrical wavelength of the alternation of the wiggling magnetic fields until the predicted wavelength is in the X-ray region?

There seems little doubt that lasers working on this principle will be designed and built in the years ahead. For the time being, however, the wavelengths at which free-electron lasers are being designed or played with are modestly long. The first experiments at Stanford University demonstrated laser action at 10.6 μm , well into the infrared. Although there are more conventional types of lasers which function in this region of the spectrum, the prospect that the wavelength of radiation from a free-electron laser may be adjusted simply by changing the energy of the electron beam suggests that devices built on such a principle should be more 'tuneable' than more conventional infrared lasers. The immediate prize in the sights of those now developing this new technology is, however, that of being able to construct lasers that operate in the ultraviolet. The potential benefits of such devices are self-evident. For the first time, it would, for example, be possible to make precise studies of several photochemical processes at present inaccessible to observation. The snags are unfortunately also clear. Free-electron lasers are likely to be less efficient at increasing energies (although much depends on the geometry) while technical problems of arranging for a geometrically rapid corrugation of a magnetic field entail the complications of handling superconducting magnets. This part of the game, nevertheless, may be well worth the candle because of the potential usefulness of ultraviolet lasers in the separation of uranium isotopes or the stimulation of thermonuclear fusion.

The potential of free-electron lasers seems to have been clearly recognized in the United States, where much technical development is under way. The best known plan is that of the Brookhaven National Laboratory on Long Island, where a 700-MeV storage ring will be commissioned next year with the intention of producing laser radiation in the optical and near-ultraviolet in 1982. Most other projects, in which industrial companies such as TRW Incorporated and Bell Telephone Laboratories have taken an interest, appear to be aimed at exploring still inaccessible parts of the infrared spectrum. Given the present limitations of even superconducting magnet design, it is inevitable that people should be forced to use beams of electrons with energies approaching 1 or more GeV. But these, of course, are early days. With any luck, free-electron lasers should be operating in the ultraviolet within the next few years. There after, it should be relatively easy to decide how the technology might be pushed still further. By the end of the decade, soft X-ray laser sources, however inefficient, should be in operation. The hard X-ray source that will make possible direct measurement of atomic and molecular energy levels will be further ahead, but not out of reach. □

Gene regulation: new, old and remote controls

from Adrian Minty and Peter Newmark

A few months ago in these columns, due warning was given that the 5' ends of every gene under the sun were "in for a beating with mutagens and nucleases in the quest for the eukaryotic promoter" (*Nature* **285**, 356; 1980). Two European meetings* with many presentations in common have since demonstrated the validity of that prediction, whilst illustrating that introns and regions far removed from the 5' ends of genes (see Fig. 1) are also coming under close scrutiny as possible sites of gene regulation.

The Hogness/Goldberg/TATA box

It has been recognized for some time that most of the eukaryotic genes which are transcribed by RNA polymerase II have a very similar, short nucleotide sequence about 30 nucleotides upstream from the place where transcription begins. The sequence is invariably AT rich and is most commonly known as the TATA (tar-tar) box but at other times as the Hogness box, since it was first recognized in the Stanford laboratory of D. Hogness. Purists, however, note that the only formal citation for the discovery at present is to the 1979 Stanford University thesis of M. Goldberg and hence refer to the Goldberg-Hogness box.

Is the TATA box essential for the initiation of transcription? An answer of sorts is emerging from a comparison of the *in vitro* and *in vivo* transcription of intact, cloned genes with cloned genes that have mutated, deleted and/or replaced TATA boxes.

For *in vitro* transcription there is, at first sight, a good deal of evidence that the TATA box is essential. Most strikingly, mutation of the chicken conalbumin gene box from TATA to TAGA reduces

transcription by more than 95 per cent (P. Chambon *s*, Strasbourg). In cruder tests, the elimination of TATA box regions of the chicken ovalbumin gene (M. J. Tsai *a*, Houston), the β globin gene (R. Flavell *a*, National Institute for Medical Research, London) or the adenovirus late region (P. Sharp *s*, MIT) reduced or halted the initiation of transcription.

Unfortunately, matters are more complicated than that. For example, several viral genes do not have TATA boxes. Furthermore, both R. Kamen (*s*, ICRF), working with the polyoma virus early region, and P. Chambon, working with the SV40 early region, reported that deletion of the TATA box regions did not prevent transcription but did shift the position(s) at which it started. From TATA box deletions and a series of other deletions, Chambon is of the view that the efficiency of transcription is controlled both by the TATA box and by regions lying downstream from it, close to the site of initiation of transcription, while the TATA box alone is responsible for fixing the site of initiation. Not everyone's data seem to fit even that fairly unspecific view and that is the way matters are likely to stay until a good deal of progress has been made in defining the ideal conditions for *in vitro* transcription.

Meanwhile, a complementary approach is to look at *in vivo* transcription. Here there is general agreement that deletion of the TATA box affects the site far more than the efficiency of initiation. This is so for the SV40 early region in monkey cells (Chambon *s*), sea urchin H2A histone genes in *Xenopus* oocytes (M. Birnstein *a*, Zurich), yeast iso1-cytochrome *c* genes in yeast (G. Faye *s*, Seattle and Orsay) and mouse β globin genes in monkey cells (D. Hamer *s*, NIH).

Upstream controlling region

Much the most interesting and broadly consistent result to come, in recent months,

from a spectrum of *in vivo* studies, is that there are regions further upstream than the TATA box which can modulate RNA polymerase II-dependent transcription. What is more, whereas deletion of these upstream controlling regions can drastically reduce *in vivo* transcription, it can be without effect on *in vitro* systems.

For example, P. Dierks (Institute of Molecular Biology, Zurich) reported that deletion of the CAAT box, found 80 nucleotides upstream from the cap site of a number of globin genes, markedly reduced their transcription in mouse L-cells, although he agreed with Flavell that this deletion was ineffective *in vitro*. In the case of the SV40 early region, Chambon reported that the crucial upstream region lay within a GC-rich region of a 72-base pair repeated sequence that, at its closest, is some 100 base pairs upstream from the TATA box.

Another interesting example is the histone H2A gene (C. Hentschel *s*, Zurich). Deletion of a segment of DNA positioned from 184 to about 520 base pairs upstream of the initiation site reduced transcription by 15–20-fold. Mysteriously, the inversion of this region, which is characterized by its AT-richness and some largish palindromes, led to a fivefold increase in transcription.

As a final example of remote control of gene activity, there are the wild-type strains of *Drosophila* which do not show the normal 'puffing' of band 3C11-12 on the X chromosome in response to ecdysone. D. Hogness (*s*, Stanford) reported that the relevant gene in each of four such strains had a 50–150-base pair deletion approximately 200 base pairs upstream from the 5' end.

How does the remote control of transcription work? There are at least three possibilities. The first is that the upstream region binds a protein which affects transcription. The second is that it binds

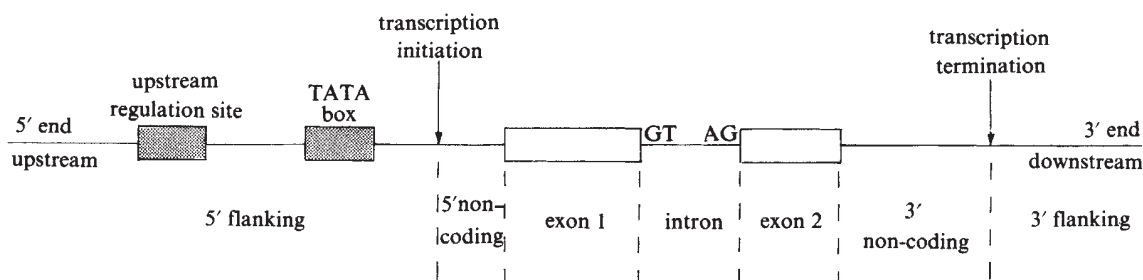


Fig. 1 Consensus polymerase II gene structure. Not to scale.

*Fourth Arolla Workshop, Arolla, Switzerland, 18–22 August 1980 and INSERM conference on 'Structure and Expression of Eukaryotic Genomes', Domaine de Seillac, 21–25 September 1980. Speakers at the former are identified by an *a* before their affiliation and those at the latter by an *s*.

RNA polymerase II, secondary structure being invoked to explain how the enzyme also binds to sites near to the point of initiation. And the third, and most fashionable, explanation is that the upstream sites phase (determine the spacing of) the nucleosomes in relation to the DNA sequence, the phasing determining whether or not a downstream gene is transcribed.

That chromatin structure plays a part is suggested by the hypersensitivity to DNase I of the 5' ends of the heat-shock genes (S. Elgin *a*, Harvard; see also Wu, C. *Nature* **285**, 854; 1980) and histone genes (A. Worcel *a*, Princeton) of *Drosophila*. Results that fit that trend are also beginning to emerge from a study of the DNase I hypersensitivity of SV40 in infective or integrated form (C. Cremisi *s*, Fred Hutchinson Cancer Center, Seattle). Furthermore, electron microscopy shows there to be a break in the regular dispersal of nucleosomes around the SV40 minichromosome close to the presumed promoter site (Y. Aloni *a*, Weizmann Institute and *Nature* **285**, 263; 1980; M. Yaniv *s*, Pasteur Institute and *Cell* **20**, 65; 1980). Polyoma virus, which has a similar nucleosome-free region, does not normally grow on undifferentiated teratocarcinoma cells. Yaniv, however, described a mutant which could grow on those cells and which had an altered DNA sequence in the presumed promoter region and a displaced nucleosome-free region, reinforcing the view that there is a relationship between the sequence of DNA, its organization into nucleosomes and the control of transcription.

Introns

As more introns are sequenced, new features of interest emerge. These include the recognition of repetitive sequences within the introns of the ovalbumin-related 'X' gene (J.-P. Mandel *s*, Strasbourg) and the vitellogenin gene (G. Ryffel *a*, University of Bern).

It is customary to believe that introns are spliced out whole, from the GT at the 5' end to the AG at the 3' end (the junction or boundary sequences). But E. Avvedimento (*s*, NIH; see *Cell* **21**, 689; 1980) has identified three possible internal splice sites in an intron of the chick α 2-collagen (type 1) gene. He proposed that a first step in the removal of this intron could be the splicing of the 3' end to any one of the three internal sites and backed up that proposal with evidence of the presence of the predicted RNA species in the nucleus of chick fibroblasts. A two-step splicing out of one of the introns of a mouse α -amylase gene is similarly indicated (O. Hagenbüchle and U. Schibler *a* and *s*, ISREC, Lausanne).

A more interesting feature of the α -amylase system is that it may provide the first example in eukaryotes higher than viruses of the production of mRNAs with alternative 5' ends from the same DNA sequence. That at least is the explanation favoured by the Lausanne group for the

fact that the major species of α -amylase mRNA in the mouse liver is identical to that in the salivary gland except that its 5' non-coding sequence is extended by about 100 nucleotides. Analysis of genomic clones from the two tissues supports the view that the alternative mRNAs are derived by tissue-specific transcription and splicing of an identical DNA sequence. Why and how this happens is uncertain.

It has been postulated that splicing is initiated by the base-pairing of one of the small nuclear (Sn) RNAs, called U1, to the intron-exon boundary sequences in gene transcripts. M. Jacob (*a*, Strasbourg) presented a list of 60 intron-exon boundary sequences in eukaryotic pre-mRNAs and concluded that, in the majority of cases, the homology between these and U1 was probably not, in itself, sufficient to ensure a stable hybrid. Approximately 40 per cent of SnRNAs in the cell are associated with heterogeneous nuclear RNA at a ratio of 1 or 2 molecules of SnRNA per 2,500 nucleotides, indicating that they are not structural components of heterogeneous nuclear ribonucleoproteins (Jacob). T. Pederson (*a*, Worcester Foundation) reported from cross-linking studies *in vivo*, that U1 is actually bound to heterogeneous nuclear RNA in the cell. This appears not to be the case for another 'processing RNA' candidate, the adenovirus VA RNA, which is not found in adenovirus nuclear ribonucleoproteins (Jacob *a*; J.M. Blanchard, Montpellier).

A different mediator of splicing was proposed by P. Slonimski (*a*, *s*, Gif) for the transcript of the yeast mitochondrial gene for cytochrome *b*. He proposes that certain introns are spliced out with the help of proteins ('maturases') encoded in the introns themselves.

Several facts favour that interpretation. They include the recent discovery that the second intron of the cytochrome *b* gene contains an open reading frame of considerable length while mutants containing stop codons within that intron accumulate unspliced transcripts of the gene, a defect that can be corrected by *in vivo* complementation. Furthermore,

those mutants contain proteins of the right size to be the 'maturase' truncated in accordance with the position of the stop codon. Neat though this system is, particularly in terms of autoregulation, there is very little evidence for its generality.

Post-transcriptional regulation

That gene expression can be regulated after, as well as at, transcription was made clear by several contributors. M. Rosbash (*a*, Rosenstiel Center, Waltham) described a temperature-sensitive mutant of yeast (RNA 2) in which the synthesis of ribosomal protein is inhibited despite normal transcription of the ribosomal protein gene. Since the unspliced transcript accumulates in the nucleus, it seems probable that the primary defect lies in the processing of the transcript to mRNA. A similar situation exists for globin mRNA in erythroblasts transformed by avian erythroblastosis virus (A. Therwath *a*, IRBM). J. Nevins (*a*, Rockefeller, New York) showed evidence that a part of the adenovirus genome expressed late in infection is also efficiently transcribed early in infection but that few cytoplasmic mRNA molecules are formed from the transcripts; again this reflects a defect in processing since polyadenylated precursor molecules accumulate in the nucleus. However, the effects of inhibitors of protein synthesis on adenovirus expression led L. Phillipson (*a*, University of Uppsala) to propose that a viral gene product encoded in the early region controls, either directly or indirectly, translation of early adenovirus mRNA *in vivo*.

Finally, Hereford (*a*, Rosenstiel Center) reported that duplicating the yeast histone genes by genetic manipulation resulted in a doubling of transcription whilst the overall level of histone mRNA remained constant due to a halving of the mRNA half-life. Further elucidation, at a molecular level, of the control of eukaryotic mRNA production remains a major task for the 1980s. □

Adrian Minty is a research fellow at the Pasteur Institute. Peter Newmark is Deputy Editor of Nature.

Deep drilling in African lakes

from N.J. Shackleton

THE climatic environmental history of Africa is a major unknown on a time scale from thousands to millions of years. So far as the Pleistocene is concerned, we are building up a fairly complete picture of the characteristics of climatic change, both on the northern continents and over the oceans. Moreover, we have sufficient information to relate reliably the time sequences from the oceans with those of

glaciated areas. To a first approximation, climate changed more or less synchronously in North America, Europe and over much of the sea surface; however, the picture of climatic change in Africa has been surprisingly different.

In the past, the dry areas of Africa have experienced episodes of much greater wetness, which gave rise to increased lake areas, bigger rivers and more extensive

vegetation. At first it was assumed that these 'pluvial episodes' occurred at times of glaciation in the north. However, the most recent 'pluvial' is reliably dated over large areas of Africa to about between 12,000 and 5,000 years ago. Unfortunately, our records from Africa are too young to allow investigation of the previous occurrence of similar pluvial episodes. Thus we cannot usefully integrate the history of tropical rainfall with older records of other climatic variables. Indeed, we have no idea of the frequency of pluvials during the Pleistocene.

Between 21 and 27 August, an international working group, convened by Dr Dan Livingstone (Duke University, North Carolina) and Dr Neil Opdyke (Lamont Doherty Geological Observatory), met at the offices of the National Science Foundation (NSF) in Washington to plan a major drilling project to tackle this and related problems. Funds for the meeting came from the Anthropology Division of the NSF, but the disciplines represented were numerous, including palynology, magnetostratigraphy, organic and inorganic geochemistry, and the study of diatoms, stable isotopes and evolutionary biology as well as African prehistory. As well as conventional dry-land drilling expertise, we had a representative of the

Deep-Sea Drilling Project whose presence was appreciated partly because of the potential value of their drilling experience (especially with the new Hydraulic Piston Corer) and partly because of their unique experience in managing a major sediment-sampling programme for the benefit of a very large scientific community.

The NSF has already funded a seismic survey which will be conducted during the next year and which will hopefully cover three lakes. A great deal of the meeting was devoted to a discussion of the relative scientific and logistical merits of over a dozen tropical lakes in Africa (we resisted the temptation to spend much time on lakes which seem logistically impossible at present). The lake most suitable for a first sampling is Lake Bosomtwe in Ghana. This occupies a crater that is thought to have been formed by meteoric impact about a million years ago. Preliminary study of short cores covering the past 30,000 years or so shows that it preserves a good vegetational history and, so far as the million years recorded is concerned, is the most attractive of the lakes we discussed. Drilling of this lake should recover very valuable material for studying the early diagenesis of organic-rich sediments. The hypothetical origin of the Ivory Coast tectite-strewn field will be tested at source.

Most important, we will have a datable and continuous record of the climate of the past million years in the tropics where rain, rather than temperature, is the most important climatic variable. The extent to which this record can be explained in terms of current hypotheses will be an important test of our present understanding of the causes of climatic change.

With regard to the potential for a record spanning many million years, Lakes Malawi, Tanganyika and Turkana seem to offer the greatest number of advantages and should be covered in the seismic survey effort.

It was clear to the meeting delegates that the environmental history of any of the East African lakes must, over the long term, be intimately connected with its tectonic evolution, and conversely, that any deep drilling in one of these will arouse considerable interest from a wider section of the geological community. The consensus of the meeting was that a project to drill in the major African lakes, although expensive, will certainly be worthwhile in terms of the interest the results will hold for many different disciplines. □

N.J. Shackleton is at the Sub-department of Quaternary Research, University of Cambridge, UK.

Herpes simplex virus in latent infection

from Howard Marsden

FOLLOWING initial infection of humans, herpes simplex virus (HSV) appears to be maintained in a latent state either for the lifetime of the individual¹ or in some individuals is periodically reactivated to produce overt lesions. Important questions arise, such as: in what tissue(s) is the virus latent, and what is the molecular basis of latency?

Goodpasture's suggestion² that HSV is latent in neuronal tissue has been substantiated in the mouse³ where, after experimental infection in the footpad, two phases could be distinguished. First is an acute phase of one or two weeks during which virus can be recovered from cell-free homogenates of lumbrosacral ganglia. This is followed by a latent phase during which infectious virus is no longer detectable but in some animals virus can be recovered by co-cultivating explanted ganglia with sensitive cell monolayers on which the virus can form plaques. In this way, HSV-1 could be recovered not only from the peripheral but also, less frequently, from the central nervous system, though not from any non-neural tissue⁴. This observation suggested that in

the mouse, HSV latency is restricted to tissues of neural origin. However, Scriba subsequently demonstrated⁵ that HSV-2 was recovered preferentially, and HSV-1 exclusively, from the footpad (the site of primary inoculation) of latently infected guinea pigs. More recently, Hill, Harbor and Blythe similarly recovered HSV-1 from the skin of latently infected mice⁶, although earlier attempts by other investigators had been unsuccessful. Are these really conflicting observations or do they reflect differences in the efficiency with which virus can be detected?

Latency involves three separate but related processes — not only establishment and maintenance but also reactivation. The virus assay method by nucleic acid hybridization used by Cabrera *et al.* (*Nature* this issue, p.288) does not necessitate formation of infectious virus particles and thus allows discrimination between the establishment and maintenance steps and the reactivation step. During the acute stage following corneal inoculation of HSV, infectious virus could be detected both in brain tissue (90 per cent of mice positive) and

trigeminal ganglia (100 per cent of mice positive); 8 weeks after inoculation, during the latent phase, virus could be reactivated by explantation and co-cultivation from 95 per cent of the trigeminal ganglia but from only 5 per cent of the brain tissue explants — yet by DNA reassociation techniques, HSV DNA sequences were detected in 30 per cent of brains of mice harbouring latent HSV in their trigeminal ganglia. The authors conclude that virus is not eliminated from the brain tissue but is maintained in a state which cannot be reactivated by the explantation techniques used.

What is this state and are all sequences of the HSV genome present? Inability to reactivate in the absence of complete genomic sequences would hardly be surprising. However, the presence and expression of an incomplete genome could have considerable consequences for the organism. Perhaps nucleic acid hybridization studies using cloned fragments of the genome will provide some insight.

Another approach to detect HSV genetic material in explant organ cultures is to superinfect explanted human ganglion

tissue with temperature-sensitive (*ts*) mutants of HSV-1 (ref. 7). Using this method, Brown *et al.* recovered wild-type virus from explant cultures which had been negative by co-cultivation for 45 days, and inferred that the *ts* mutant must have been complemented or rescued by genetic information in the ganglia. The authors suggest use of restriction enzymes to try to detect in the putative rescued virus DNA sequences which are not present in the *ts* mutant.

An important question is whether any of the viral genome is expressed during the latent state. Puga *et al.*⁸ could find no HSV mRNA using reassociation techniques capable of detecting one genome equivalent of mRNA in 2,000 cells. On the other hand, using *in situ* hybridization, Galloway *et al.*⁹ reported HSV mRNA in 0.4–8 per cent of human ganglia. HSV-specific thymidine kinase (TK) activity was observed in the sensory ganglia of mice up to 60 days after infection¹⁰ but not in the skin or ganglia of guinea pigs, even with an assay which could detect the activity of one lytically infected cell in 3,000 uninfected cells¹¹. We do not yet know whether any of the observed expressions are necessary for maintenance of the latent state or whether they reflect the spontaneous reactivation of cells which may subsequently be eliminated from the animal.

A different approach with potential for recognition of the viral functions necessary for latency is the use of HSV mutants. Tenser and colleagues^{12,13} found that TK-negative mutants of HSV established latent infections in the trigeminal ganglia of both guinea pigs and mice with a lower frequency than that of TK-positive virus, suggesting that expression of this enzyme is necessary for latency. A collaborative study between Stevens' group (University of California, Los Angeles) and Subak-Sharpe's (Glasgow, UK) identified six temperature-sensitive mutants of HSV-1 which produce latent infections in mice with reduced frequency. One of these mutants (*tsK*) has a lesion in the immediate-early polypeptide V_{mw} 175 (ref. 14) and appears to be blocked at or near the immediate-early stage of infection. The combined results of Preston¹⁴ and those of the mutant latency study suggest that expression of at least one immediate-early function and one or more later virus functions is necessary for latency¹⁵.

The techniques hitherto employed in the isolation of *ts* mutants generate mutations in genes essential for growth in tissue culture. Genetic manipulation techniques will also allow isolation of mutants lacking sequences not essential for growth in tissue culture but which might be essential for latency. Screening of such deletion mutants for latency should be informative. □

Howard Marsden is at the MRC Virology Unit, Glasgow, UK.

1. Barringer, J.R. & Swoveland, P. *New Engl. J. Med.* **288**, 648 (1973).
2. Goodpasture, E.W. *Medicine* **8**, 233 (1929).
3. Stevens, J.G. & Cook, M.L. *Science* **173**, 843 (1971).
4. Cook, M.L. & Stevens, J.G. *J. gen. Virol.* **31**, 75 (1976).
5. Scriba, M. *Nature* **267**, 529 (1977).
6. Hill, T.J., Harbour, D.A. & Blyth, W.A. *J. gen. Virol.* **47**, 205 (1980).
7. Brown, S.M., Subak-Sharpe, J.H., Warren, K.G., Wroblewska, Z. & Koprowski, H. *Proc. natn. Acad. Sci. U.S.A.* **76**, 2364 (1979).
8. Puga, A., Rosenthal, J.D., Openshaw, H. & Notkins, A.L.

- Virology* **89**, 102 (1978).
9. Galloway, D.A., Fenoglio, C., Shevchuk, M. & McDougall, J.K. *Virology* **95**, 265 (1979).
10. Yamanoto, J., Walz, M.A. & Notkins, A.L. *Virology* **76**, 866 (1977).
11. Fong, B.S. & Scriba, M. *J. Virol.* **34**, 644 (1980).
12. Tenser, R.B. & Dunstan, M.E. *Virology* **99**, 417 (1979).
13. Tenser, R.B., Miller, R.L. & Rapp, F. *Science* **205**, 915 (1979).
14. Preston, C.M. *J. Virol.* **32**, 357 (1979).
15. Watson, K., Stevens, J.G., Cook, M.L. & Subak-Sharpe, J.H. *J. gen. Virol.* **49**, 149 (1980).

New directions in spectroscopy

from Derek Stacey

ATOMIC SPECTROSCOPY has a long and distinguished tradition. Its history is closely linked with that of quantum mechanics; atomic structure and the interaction of atoms with radiation provided both the inspiration and the testing ground for many of the theoretical advances which took place in the first half of this century, from the early quantum theory of Bohr to quantum electrodynamics (QED). Now that the emphasis has been passed on to elementary-particle physics, what remains to be done in atomic spectroscopy? A recent conference* showed how spectroscopists have responded to this challenge.

First, there are some areas in which fundamental questions can still be studied. One paper, given by a team from CERN and the University of Pisa, described progress in an experiment to measure the birefringence of empty space (E. Iacopini and E. Zavattini *Phys. Lett.* **85B**, 151; 1979). According to QED, ordinary plane-polarized light becomes slightly elliptically polarized when travelling through an evacuated region in which a magnetic field exists. As there is nothing else present, it follows that the photon-photon interaction itself is responsible for the birefringence. Measurement of the induced ellipticity would give a direct test of this interaction at very low energies; this would be interesting because the birefringence is a genuine vacuum-polarization effect, arising only as a result of virtual pair creation. Apart from the interest in testing QED, the effect could have implications for astrophysics when one is dealing with the propagation of light through regions which may be subject to strong magnetic fields.

The problem with the experiment is the smallness of the effect. The idea is to shine a linearly polarized beam from an argon-ion laser into an evacuated region where there is a magnetic field of 8 teslas, provided by superconducting magnets; the effect is proportional to the fourth power of the field, so it is important to make this as large as possible. A mirror system causes the beam to cross this region many times, so that the effective path length through

the field is several kilometres. Even so, the ellipticity predicted for the emerging beam is only around 10^{-12} . The measurement is carried out using familiar methods of analysing polarized light, the object being to detect changes which occur synchronously with the switching of the magnetic field.

The project brings to mind the experiments which use rather similar optical techniques to measure the rotation of the plane of polarization of light when it passes through an atomic vapour; the rotation gives a measure of parity non-conserving effects in the atomic system and hence demonstrates the presence of weak, neutral currents in atoms. These experiments are of fundamental interest and considerable effort has been put into them, but they have still taken several years to measure effects at the 10^{-7} level. Thus the prospects for the new test of QED seem discouraging at first sight, but this could be a misleading comparison. Much of the difficulty of the parity experiments arises because one has to search for an effect which is sensitively dependent on wavelength. This problem does not arise in the QED test, in which the switching of the magnetic field provides the only necessary discriminating factor. Nevertheless, success in this experiment would be a major technical achievement.

Although comparatively few groups are working on such fundamental problems, the contributed papers showed that the field of atomic spectroscopy, far from declining, is showing a remarkable resurgence. This is largely due to the profound influence that the tuneable dye laser has had on atomic physics. During its early days, powerful spectroscopic techniques based on the device were developed but were usually applied to systems already well understood. Only now is the full potential of dye-laser methods becoming apparent; they have transformed established fields of study and opened entirely new areas for investigation — Rydberg states, collisional redistribution of radiation, super-radiance and so on. The attraction of this type of work is easy to see; in many branches of physics, anyone wishing to take advantage of advances in technology must join a team working at a major central faculty. The tuneable laser is a comparatively cheap

*Annual conference of the European Group for Atomic Spectroscopy, held in Pisa, 2–5 September 1980.

device which permits a whole range of new experiments to be carried out in an ordinary laboratory.

Several papers described improvements in dye-laser techniques, but the most significant development in this area has been the emergence of the new generation of dye lasers based on the ring geometry. Ring lasers are inherently much more powerful and stable than the earlier standing-wave devices, and one reason for their importance is that they allow tuneable-laser methods to be applied in the ultra-violet. Most atoms have hitherto been inaccessible to the simpler dye-laser techniques because the tuning range of available dyes does not extend to short enough wavelengths. A ring laser operates with the same dyes but gives sufficient power to allow frequency doubling, or frequency generation, by mixing with the output of fixed-frequency lasers.

Another reason for the present interest in atomic spectroscopy is that the various techniques developed over the years primarily as spectroscopic tools are finding wider application. An invited paper presented at the conference by F. Laløe (Saclay Nuclear Research Centre, Paris) dealt with the properties of polarized ^3He (that is, nuclear spins all parallel) (C. Lhuillier and F. Laløe *J. Phys.* **40**, 239;

1979). The connection with atomic spectroscopy is that the most promising of the methods proposed to prepare a sample of this material is based on the standard spectroscopic technique of optical pumping. (In particular, a newly explored colour centre in NaF seems well suited for pumping the $2^3\text{--}2^3\text{P}$ transition, an important step in the preparation process [L.F. Mollenauer *Opt. Lett.* **5**, 188; 1980].) The talk reviewed the remarkable properties predicted for this fluid; they derive from the indistinguishability of the particles and can give rise to 'quantum effects' even when the system is in the form of a low-density gas. This is because the transport properties depend on the collision cross-section, and the Pauli Principle prevents the atoms from hitting each other. Other intriguing predictions include changes in the liquid-vapour and liquid-solid equilibrium pressures (C. Lhuillier and F. Laløe, *op. cit.*). It may be some time before any of these effects are demonstrated practically. However, the project is a good example of the way in which atomic spectroscopy continues to influence the development of other areas of physics. □

Derek Stacey is at the Clarendon Laboratory, University of Oxford, UK.

Precambrian evolution

from Thomas D. Brock

THE PRECAMBRIAN constitutes the longest but least understood period of Earth history. In terms of biology, it is the period during which life arose and underwent extensive biochemical diversification. In terms of geology, it is when the continents and oceans became segregated and defined. Although today life plays a crucial role in geochemical processes at the surface of the Earth, geochemical events during at least part of the early Precambrian probably occurred in the absence of biological catalysis.

A number of ancient sedimentary deposits, some with important ore-bearing rocks, are accessible for study and analysis, the oldest of which is the Isua Supracrustal Group in Greenland (around 3.8×10^9 years before the present [BP]). The only real clues to evolutionary processes in the Precambrian are found in the chemical and biological constituents present in these ancient rocks, and the aim of students of the Precambrian is to try to make sense out of the fragmentary records available. Although this is interdisciplinary research at its highest level, chemists, geologists and biologists rarely have the opportunity of meeting and working with one another. It was therefore particularly appropriate that a conference* on Precambrian evolution was arranged that involved both geochemists and bioscientists.

Four separate topics were considered, each by a separate group containing chemists, geologists and biologists: (1) banded-iron formations; (2) stratified sulphide deposits; (3) organic geochemistry and palaeomicrobiology of ancient sediments; and (4) biogeochemical evolution of the ocean/atmosphere system. Although these topics were considered by separate groups, they so greatly overlapped that it was necessary for individuals to move from one group to another from time to time, a process encouraged by the format of the Dahlem Conferences.

Of over-riding importance for an understanding of Earth history is the estimation of the sequence of events leading to the formation of an oxygen-containing atmosphere. Oxygen arises today almost exclusively through plant photosynthesis, and the single most important feature in the evolution of higher-life forms was the appearance of this process. This must have occurred early in Earth history, but we do not know precisely when. Data from studies on chemical evolution suggest that life probably arose in an initially anoxic environment, and micropalaeontological evidence is strong that, by the Cambrian, the Earth's atmosphere contained nearly as much oxygen as it does today.

The geochemist looks carefully through

the rock record for minerals or chemical constituents that require oxygen for their formation. Unfortunately, although they exist, the time of their formation in the geological record fails to provide unequivocal evidence for an oxygenic atmosphere, and one major point to come out of the conference was that no single piece of evidence, in itself, would be sufficient to establish the existence of an oxic atmosphere. Rather, a whole sequence of analyses must be done on the same rocks in order for valid interpretations to be possible. Analyses have so far been too fragmentary to provide firm conclusions, although the weight of opinion, based primarily on the microfossil record and on iron and uranium mineralogy, is that oxygen probably first arose in significant quantities around 2 to 2.6 billion years BP. However, the interpretation of microfossil evidence is fraught with difficulty, as there are no microbial structures that can be unequivocally associated with oxygen evolution.

One of the four groups evaluated the evidence concerning Precambrian banded-iron formation. These formations are extremely important because they provide over 90 per cent of the minable iron deposits of the world. There was a major period in Earth history between 2 and 2.6 billion years BP when iron was precipitated in vast amounts, and after this time only very small deposits of this type of iron have deposited. The main Precambrian iron formation is rich in silica (chert) and has extensive characteristic banding (macro-, meso- and micro-bands), thus prompting the name banded-iron formations (BIF). These deposits probably formed as a result of sedimentation of oxidized iron on to ocean shelves, and because of the known chemistry of iron, it is clear that the Precambrian oceans could only have held in solution the vast amounts of iron necessary for BIF if the ocean waters were much less oxygenated than they are today.

The favoured model involves a circulating Precambrian ocean with an upper oxygenated layer and a lower anaerobic zone, the latter of which held a vast amount of ferrous iron. Movement of reduced iron into the surface waters, perhaps as a result of oceanic upwelling (analogous to the movements which bring up deep, cold, nutrient-rich waters on to shallow shelves in the present oceans), brought soluble iron to surface waters from where it was probably carried into closed basins for precipitation and sedimentation. The upwelling analogy involving marginal basins is strengthened by the facts that iron formations are enriched in phosphorus,

*The Dahlem Conference on Mineral Deposits and the Evolution of the Biosphere, held during the first week of September 1980. A workshop report is also available: H.D. Holland and M. Schidlowski (eds) *Mineral Deposits and the Evolution of the Biosphere*, Dahlem Konferenzen (Verlag Chemie, Weinheim, FRG, in the press).

and by the purity of these deposits (absence of heavy metals). It is thought that the world-wide appearance of BIF between 2 and 2.6 billion years BP may be largely due to the development, for the first time, of wide continental shelves and enclosed basins adjacent to the deeper, oceanic areas. The cyclicity which led to banding was probably the result of seasonal climatic changes, perhaps controlling phytoplankton oxygen production.

During the discussion, it became clear that considerably more information is needed on iron oxidation states in Precambrian iron formations, as an indication of the oxidation states of the Precambrian oceanic waters. Since the presence of iron in Precambrian rocks provides some of the best evidence for the oxygen status of the atmosphere, it is essential that the oxidation state of iron during primary precipitation be established with certainty. Does the massive development of iron formation at 2 to 2.6 billion years BP represent the initial rise in oxygen in the atmosphere, or is it more related to the initial development of continental shelves, providing an appropriate habitat for iron deposition?

Stratified sulphide deposits display a host of features indicating that they formed at or near the sediment-water interface. Although it had been considered previously that some of these deposits were biogenic, the conclusion of the Dahlem workshop was that they had to have an external (possibly exhalative) source of metal and that the sulphur may represent a mixing of multiple sources (biogenic and hydrothermal).

An important biogeochemical tool, sulphur isotope fractionation, was carefully re-examined in the light of the evidence. Two components are needed for the formation of a sulphide deposit — a source of sulphide and a source of iron (to tie up the sulphide). Sulphide can come from either abiotic sources (hydrothermal or volcanogenic) or biotic sources (sulphate-reducing bacteria). Because bacterial sulphate reduction leads to the discrimination of ^{34}S against ^{32}S , it has been conventional to use isotope ratios of sulphur in sulphide deposits to infer biogenicity. However, more recent work on the isotope ratio distributions in sedimentary and hydrothermal deposits shows that a clear-cut conclusion on biogenicity is not possible. Most ancient sulphide deposits have a high iron content, and it was shown that such a high iron content is unlikely to develop in a typical modern sedimentary setting where biogenic sulphide is depositing. Modern analogues for the ancient stratified sulphide deposits are the Red Sea brine pools and the 'black smokers' at 21°N on the East Pacific Rise. If the Precambrian sulphide deposits are indeed largely abiotic, then it is impossible to determine

Thomas D. Brock is at the University of Wisconsin.

from an examination of these deposits the time of onset of biological sulphate reduction. It was concluded that extensive new data will be needed to elucidate the onset of this important step in biochemical evolution.

Many Precambrian deposits have extensive amounts of reduced carbon, analysis of which may provide clues to biological evolution. However, interpretation of such analyses requires careful application of concepts developed during studies on modern ecosystems. There is now available a wide variety of techniques that could be used in the study of Precambrian organic matter, including ion microprobe, pyrolysis-gas chromatography, high-resolution mass spectrometry (MS), photoacoustic spectroscopy, laser pyrolysis, high-pressure liquid chromatography (HPLC), HPLC-MS, high-resolution capillary gas chromatography, ^{13}C nuclear magnetic resonance and X-ray diffraction. It is clear that a vast amount of information can be obtained, but it is vital that the analyses be carried out on carefully collected specimens, and that, as far as possible, combinations of methods be used, to permit cross-checking and a more precise interpretation.

Organic matter in ancient rocks can be divided into two major fractions, extractable organic matter (EOM) and kerogen (nonextractable). Because of the ease of analysis, EOM has been studied more extensively, but it is essential that the EOM studied be syngenetic with the rock and not a post-depositional replacement or contamination arising during processing.

The interpretation of the chemistry of both EOM and kerogen requires a knowledge of the organic chemistry of modern ecosystems and of how organic matter changes during diagenesis and catagenesis. Micropalaeontological studies are also vital as they provide clear evidence that life was involved in organic matter deposition or modification. Studies on carbon isotopes in ancient organic matter may provide evidence for biogenicity, provided the data can be interpreted properly in relation to modern processes (although the example for sulphur isotopes discussed above shows the dangers involved).

During the conference, it was frequently the case that biologists were turning to chemists for appropriate chemical evidence that would aid their studies, and that geologists were seeking biological evidence to confirm or expand their conclusions. One of the main conclusions of this conference was that no one science alone would provide an answer, and that progress can be made only when all proceed together. Much discussion during the conference centred on the problems of communication between biologists and geologists, and the need for more formal mechanisms of interchange. Although 'biogeochemistry' has been used as a term since the mid-1930s, it is only now emerging as a discipline in its own right. Another result of the Dahlem workshop was strong consideration for the formation of a new society to be called the 'BioGeo-Chemistry Society' that would bring together workers from a variety of fields for a concerted attack on some of the most fundamental problems of our times. □

Green light in Heidelberg

from a correspondent

THE 200 or so molecular biologists who attended the sixth EMBO symposium* in Heidelberg spent 3½ days discussing gene organization and expression in green plants. The surprisingly large size of the audience might suggest that the organizers had found a patch of fertile ground and that the topic was scientifically ripe. In the event, however, there turned out to be few converts to the green revolution. For the most part, the audience consisted of those already in the field, together with a sprinkling of volunteers from other disciplines and a few weeds from the recombinant DNA companies whose presence seemed to presage a new meaning for the term 'seed-money'.

Nuclear genes

Much of the meeting was devoted to the presentation of the results of molecular cloning experiments. cDNA and genomic libraries of the genomes of several species of plants have been constructed and several genes, particularly those expressed at high levels, have been analysed in some detail. Most attention has been paid to the genes coding for the storage proteins of cereals and legumes, which are silent in vegetative tissues but extremely active at specific stages of seed development. For example, the set of genes coding for the ten or so superabundant proteins of soybeans is turned on immediately after flowering; the concentration of their mRNAs in the cotyledons increases steadily for the next 75 days, by which time 50 per cent of the total mRNA is composed of only these few species (Goldberg, University of California, Los Angeles). None or very

*Molecular Biologists look at Green Plants', European Molecular Biology Laboratory, Heidelberg, 29 September — 2 October 1980.

little of this RNA is ever detected in axis or leaf cells.

DNA copies of seven superabundant mRNAs have been made and cloned in prokaryotic vectors. By hybridization selection, *in vitro* translation and immunoprecipitation, four of them have been shown to code for polypeptides of the 7 and 11S storage proteins as well as the Kunitz trypsin inhibitor. Although these proteins appear to be coordinately synthesized, their genes are not closely clustered. Hybridization experiments between cDNAs and representative clones from a genomic library show that the genes for superabundant mRNAs occur at most a few times within the soybean genome and do not map within 20 kilobases of one another.

A major problem, therefore, is that of how small, dispersed families of apparently unrelated genes such as these are controlled and simultaneously activated. Reassociation kinetic experiments rule out the possibility that the genes are differentially amplified during embryogenesis. Furthermore, no common feature has appeared from an analysis of the gross structure of the genes that would explain their behaviour. In fact, their organization appears to be quite dissimilar — for example, the gene for Kunitz trypsin inhibitor contains an 9.7-kilobase intervening sequence, while that for 11S storage protein is apparently colinear with its mRNA. Because no nuclear transcripts of these genes can be detected in cells which do not contain storage proteins (leaf cells, for example), it seems likely that their expression is regulated primarily at the level of transcription.

The storage proteins of barley (hordein) and maize (zein) are encoded by gene families which appear to contain at least ten to twenty highly homologous members. Interestingly, the expression of the entire family is drastically influenced by the activity of other genes (Larkins, Purdue University; Feix, University of Freiburg; Brandt, Carlsberg Laboratory, Copenhagen). Thus, mutations at the *lys 3A* locus of barley and the *opaque 2* locus of maize effectively abolish expression of the respective gene families. Again, this effect seems to be mediated at the transcriptional level. Clearly it will be interesting to compare the DNA sequences which lie upstream from the structural genes coding for these various storage proteins.

Chromosomal rearrangements

Several genes coding for abundant proteins (for example, the 11S storage protein and the Kunitz trypsin inhibitor of soybeans) appear to be located in the immediate vicinity of middle-repetitive DNA (Goldberg, University of California, Los Angeles). The evolution of analogous sequences in the genome of cereals was discussed at Heidelberg by R. Flavell (Plant Breeding Institute, Cambridge). By using cloned segments of repetitive DNA as

probes in *in situ* hybridization experiments, he has shown that the greatest concentration of several species of repetitive DNA maps at the ends of cereal chromosomes.

The detailed organization of these sequences, however, varies from species to species and even from chromosome to chromosome within a single species. Thus certain repetitive sequences are present more than 10^6 times in one secale species but fewer than 50 times in another; and several examples are known where different chromosomes of a single plant species carry different repetitive sequences at their ends. All of these observations are consistent with a model in which most of the non-coding DNA of cereal chromosomes 'turns over' relatively rapidly during evolution by a process of amplification and deletion, and new sets of repetitive sequences are created by the translocation of small pieces of DNA into new sites. Why repetitive sequences in plant genomes should be sequestered to the ends of chromosomes is not known.

Chromosomal rearrangements in plants are, of course, not limited to non-coding sequences. Burr (Brookhaven) has analysed mutations at the *shrunk* locus of maize caused by the transposable element *DS* (McClintock *Yb. Carnegie Instn Wash.* 51, 212–219; 1952) which causes chromosome breaks in the endosperm leading to dicentric chromosome formation. The *shrunk* locus codes for the enzyme sucrose synthetase, a partial cDNA copy of whose mRNA has been cloned. Using this as a probe in Southern hybridizations, Burr has been able to show alterations in the structure of the *shrunk* locus as a result of *DS* action, but because of the small size of the probe the extent of these alterations is not known. At least one of them involves deletion of part of the gene while others appear to result from insertion of sizeable tracts of DNA, perhaps multiple copies of *DS* itself.

Splicing

By contrast with this plasticity, sequences coding for some plant proteins have not changed appreciably over large spans of evolutionary time: for example, the leghaemoglobins of soybeans which, judged by their amino acid sequences, appear to be related to the ancestral haemoglobin of higher animals. The leghaemoglobins are encoded by a set of six separate genes, four of which are closely linked in a single 30-kilobase DNA fragment (Marcker, University of Aarhus). This arrangement is reminiscent of those found in the haemoglobin genes of mouse and man. The leghaemoglobin *c* gene contains two intervening sequences (between nucleotides coding for amino acids 34/35 and 103/104) which map at positions almost identical with those occupied by intervening sequences in mammalian haemoglobins. A third intervening sequence, not yet recognized in

any other haemoglobin, has been detected in leghaemoglobin *c* gene and has been mapped between sequences coding for amino acids 67 and 68. Whether this is a general feature of all six leghaemoglobin genes or is specific to the one copy which by chance has been cloned and sequenced is not known.

The DNA sequences at intron–exon boundaries have also now been determined for the phaseolin gene of beans (Hall, University of Wisconsin) — it contains three intervening sequences — and for the 23S ribosomal RNA gene of *Chlamydomonas* chloroplasts, which contains one (Rochaix, University of Geneva). All of these have the dinucleotides GT and AG at their 5' and 3' ends and correspond in many other respects to the common sequence across splice junctions in heterogeneous nuclear RNA of mammalian cells. Presumably, therefore, the mechanism by which transcripts of plant genes are spliced into mature functional RNAs is fundamentally similar to that in animal cells. There are indications that small 'adaptor' RNAs may be involved in splicing of mammalian RNAs (Lerner and Steitz *Proc. natn. Acad. Sci. U.S.A.* 76, 5495; 1979). A similar set of RNAs has now been discovered in green plants. Working with *Chlamydomonas*, Rochaix has shown that the 5.85S nuclear RNAs, and the 3S and 7S RNAs encoded by the chloroplast, are partially complementary to the sequences flanking the intron within the chloroplast gene coding for 23S ribosomal RNA. By pairing with the coding sequences in the precursor RNA, these small RNAs could cause the intervening sequence to loop out, making it more accessible to the enzymatic machinery involved in splicing.

Chloroplast genes

That chloroplasts carry genes essential for plant growth and survival has been known since the early years of this century. However, it is only during the past few years that chloroplast chromosomes have been physically dissected. Restriction endonuclease mapping and electron microscopy have shown that chloroplast genomes are always circular and large in size, varying from 135–210 kilobases in different plant species. A commonly occurring feature is the presence of two large, inverted and repeated segments, which in the case of maize are 22.5 kilobases in length. Within these repeats are arranged the genes for the chloroplast ribosomal RNAs in the order 16S–spacer–23S–5S. Spacers are a constant feature of all plant chloroplast genomes so far examined but vary in size from species to species. Within the genus *Oenothera*, insertions and deletions are found to occur within them. Because such alterations always occur in both copies of the spacers simultaneously, it seems likely that a mechanism exists which corrects one copy of the spacer against the other. Why this correction process occurs is not known.

The spacer regions are neither silent nor empty, containing genes for tRNAs and, in the case of *Chlamydomonas*, for 3S and 7S RNAs as well (Rochaix). Genes for additional species of tRNA (perhaps 30 in all) are scattered throughout the repeated and unique regions of the chloroplast genome. At least some of these are split by intervening sequences, which in *Chlamydomonas*, as in yeast, are located 1–2 bases distal of the anticodon loop (Kossel, University of Freiburg).

The sequences of the 16S, 23S and 5S genes of the chloroplast of several plant species are remarkably homologous with the corresponding sequences of *Escherichia coli* (Delius, Embl; Kossel; Rochaix). These highly conserved sequences include those near the 3' end of 16S rDNA as well as those immediately flanking the intervening sequences in *Chlamydomonas* 23S RNA. It therefore appears that part of the genome of present-day chloroplasts is derived from prokaryotic precursors.

Although chloroplast and nuclear genes are carried on separate pieces of DNA, at least two lines of evidence indicate that they are not evolving independently of one another. First, within the genus *Oenothera*, where five different types of chloroplast genome can be recognized by restriction-enzyme digestion, interspecific crossing can result in combinations of nuclear and chloroplast genomes not normally found in nature (Herrmann, Botanisches Institute, Dusseldorf). Often these hybrids are extremely unhealthy, displaying a bleached phenotype. Apparently, there is a mutual dependence between the two genomes such that chloroplasts are adapted to express in only one or a few nuclear backgrounds.

Secondly, the chloroplast genome is now known to code for more than 100 polypeptides. Some of the proteins found in the chloroplast, however, are not coded by the genome of the organelle but by the nucleus of the cell. Among these polypeptides are subunits of the thylakoid-located ATP synthetase of photosystems I and II, of cytochrome *f* and of the large subunit of ribulosebiphosphate carboxylase. These examples of cooperation between the chloroplast and nucleus argue strongly that the two genomes evolve as a binary pair.

The most detailed analysis of this phenomenon has come from studies of ribulosebiphosphate carboxylase — an enzyme built from two non-identical subunits. The larger of these subunits (LS) is coded by the unique region of the chloroplast genome, the smaller (SSU) by a nuclear gene. Both genes have now been cloned (the former from maize and the latter from pea) and sequenced. The mRNA for the small subunit is 760 nucleotides in length and is derived by splicing from a nuclear precursor some 200 nucleotides longer. It codes for a protein which contains at its N-terminal end at least 13 hydrophobic amino acids which are not

found in the mature subunit. Presumably they form a signal sequence which is cleaved off as the small subunit is transported into the chloroplast.

The LS subunit of ribulosebiphosphate carboxylase occupies a special place in the history of plant molecular biology in that its gene was the first to be mapped on the chloroplast genome. The entire gene, together with its flanking sequences, has been cloned and sequenced. The coding sequence is 1.6 kilobases in length and is colinear with the polypeptide. Located on the opposite DNA strand, some 300 nucleotides away, is the *his* gene. Despite their close proximity, the two genes are expressed in entirely different ways: that for *his* is expressed constitutively in both mesophyll and the neighbouring bundle-sheath cells of maize, while that for the large subunit is expressed efficiently only in mesophyll cells. It seems quite possible that this differential control is achieved at the level of transcription, since efficient and accurate transcription of the cloned LS gene *in vitro* by DNA-dependent RNA polymerase requires a 26K polypeptide isolated specifically from chloroplasts.

Plant vectors

The relative merits were discussed of two vectors which, potentially, can replicate both in plant and bacterial cells and also allow the expression of foreign genes in plants. Both have highly unusual properties. The double-stranded DNA of cauliflower mosaic virus has recently been sequenced and has been shown to contain single-strand discontinuities at unique sites — two in one strand and one in the other (Frank *et al.* Cell 21, 285; 1980). The viral DNA propagated in bacterial plasmids and therefore without discontinuities, retains its infectivity. However, the viral genomes arising from such infections and encapsidated into progeny virus particles reacquire single-strand breaks and the tangled topology typical of the viral DNA. Thus, neither the discontinuities nor the knotted topology are necessary for establishing infection but are formed in DNA which has been replicated in plants (Howell, University of California, San Diego; Hull, John Innes Institute, Norwich).

Only the DNA strand with the single discontinuity is transcribed during infection: at early times, a single species of RNA is detected with a molecular weight equivalent to that of one entire strand of the viral genome. Later, at least four other transcripts are found, the most prominent of which is about 2,200 nucleotides in length and codes for the inclusion body protein of molecular weight 65,000. Despite intensive searching, no sites have been detected in the genome which can accept foreign sequences without affecting the viability of the virus. Until such sites are found or until a system is established in which defective viruses can be complemented, cauliflower mosaic virus cannot be used as a vector.

The second potential vector is the Ti plasmid of the Gram-negative soil bacterium *Agrobacterium tumefaciens*, which induces crown gall tumours on most dicotyledonous plants. The Ti plasmid causes the neoplastic transformation of plant cells and the resulting transformed cells contain a segment of the Ti plasmid (T-DNA) which is 12–24 kilobases long and which is integrated and transcribed (Schell, Cologne; Chilton, Saint Louis). The boundaries of the integrated T-DNA have now been cloned from a line of transformed tobacco cells and analysed by DNA sequencing and hybridization. One of the clones examined contains the right and left borders of the T-DNA linked together; another contains the right end of the T-DNA linked to repetitive plant DNA (Zambryski, University of California, San Francisco).

These data suggest that the integrated T-DNA is arranged as a tandemly repeated element within repetitive plant sequences. However, they do not shed light on the mechanism by which the integration event occurs, nor do they prove that the T-DNA can become a stable part of the plant chromosome that can be transmitted through meiosis. It is the latter property which will determine whether T-DNA can fulfill its promise of serving as vector to carry foreign genes into dicotyledonous plants.

Viroids

Of all infectious agents so far discovered, viroids must rank among the most bizarre. The best characterized — potato spindle tuber viroid — is a single-stranded circular RNA molecule 359 nucleotides in length, with so much base-pairing that it is thermodynamically equivalent to double-stranded DNA. Several strains of potato spindle tuber viroid are known which differ greatly in their capacity to cause disease. Surprisingly, these strains differ from each other in only three nucleotides (Sanger, Giessen). How such seemingly insignificant changes can result in vast alterations in biological properties is unknown.

The mechanism by which viroids replicate has long been puzzling. Too small to code for an elaborate replication machinery, it had been assumed that, like retroviruses, they may integrate into the host genome and then be transcribed into new progeny viroids. This view has been challenged by two exciting findings. First, viroids are efficient *in vitro* templates for RNA polymerase II isolated from tomato callus or leaf cells. The product is equivalent in size to full-length viroid RNA (Sanger). Secondly, viroid replication in cells is inhibited by amanitin (Sanger; Diener, Beltsville). These results open the possibility that viroids replicate by an entirely novel mechanism in which the infecting RNA molecules are copied by the host enzyme which is normally responsible for synthesis of nuclear precursors to mRNA. □

Electrochemistry and the geochemist

from H.A. Tourtelot

ELECTROCHEMISTRY deals with the relation of electron flow to chemical transformations that either create electricity or are caused by the flow of electric current. Geochemistry generally deals with Earth materials that are poorly conductive so that electrochemical transformations take place only within very short distances. In contrast, the term 'electrochemistry' seems to be used by geochemists to describe transformations involving highly conductive materials, such as sulphide minerals, and some others through which electrons travel easily. Transformations separated by considerable distances thus also are coupled by electron flow. Morris *et al.* describe in this issue of *Nature* (ref.1, p.250) such an electrochemical model for the formation of deep-seated iron ores.

Electrical phenomena around ore deposits were recognized long ago and provide the basis for the self-potential method of geophysical prospecting^{2,3}. The electrochemical mechanism for self-potentials associated with sulphide ore bodies was examined by Sato and Mooney⁴ based on the theoretical work of Pourbaix⁵ and Latimer⁶. The paper by Sato and Mooney seems to be the chief introduction to geochemists of electrochemical theory applied to ore deposits. An ore body of conducting minerals connects oxygenated ground water near the surface of the Earth with ground water at a depth deficient in oxygen. This redox gradient is 'short-circuited' by upward electron flow in the conductive ore minerals⁷. Reduced compounds at depth thus are oxidized (electrons lost) and oxidized compounds near the surface are reduced (electrons gained). The circuit is closed by ionic conduction in the pore fluids around the

ore body. Cations move upwards towards the surface and anions move downwards. The diagram shown by Morris *et al.* (*Nature* this issue, p.251, Fig.1a) illustrates the general character of electrochemical systems as well as the specific system involved in iron ores where magnetite is the conductor of electrons, ferrous iron is oxidized to ferric at depth and oxygen is reduced to OH⁻ at or near the surface.

Most economic iron ores consist of ferric oxide minerals that have formed largely by the oxidation of pre-existing ferrous minerals, such as magnetite, siderite, and iron silicates originally deposited in the Precambrian in an association known as banded-iron formation⁸. At some deposits, such as those in Brazil, formation of supergene ferric oxides that make up the ore is related to present and relatively recent weathering conditions⁹. As Morris *et al.* point out, some ferric-oxide ore deposits, such as those at Krivoy Rog, Russia and in the Hamersley Ranges, Western Australia, now exist at such great depths that their origin cannot be explained by the action of oxygenated ground waters, either now or at any time in the geological past. At Krivoy Rog, for example, oxide ores are found at depths as great as 2,400 m and in the Hamersley Ranges at depths greater than 400 m.

Explanation of the origin of the deep-seated iron oxide ores in the Hamersley Ranges by processes other than electrochemical are not offered by Morris *et al.* They suggest that electrochemical processes based on the conductivity of magnetite can be called upon to explain the origin of deep-seated oxidized banded-iron formation for ore deposits where the action of 'normal' oxidation processes cannot be comfortably accepted. Differing opinions, no doubt, will be expressed.

An earlier application of electrochemical theory to the genesis of ore deposits is that of Constantinou and Govett^{10,11}. They considered the idea that electric potential differences within a sulphide body could cause movement of metals and could affect the distribution of elements in rocks overlying ore of the Cypress sulphide deposits. Experiments by Govett and Whitehead¹² showed that sulphides of different metals develop zoning patterns that are consistent with electrochemical theory and that are similar to the arrangement of minerals in some stratiform sulphide bodies such as Rammelsberg.

Another application of electrochemical theory is in exploration geochemistry. The electrochemical model for the formation of the Hamersley deep-seated iron deposits is

H.A. Tourtelot is a geologist with the US Geological Survey, Denver, Colorado.

based on work by Thornber, one of the co-authors with Morris *et al.*, on the supergene alteration of massive nickel sulphide deposits at Kambella, Western Australia¹³. Thornber's main effort was to explain supergene alteration of sulphides, according to electrochemical principles with a view to recognizing parameters useful in geochemical prospecting. He concluded that, under favourable conditions, Eh and pH patterns could indicate the presence of an oxidizing ore body.

Govett, taking a different approach to exploration geochemistry, suggested that metal contents in soils above some mineral deposits resulted largely from the upward flow of metal cations around a conducting sulphide ore body¹¹. Later, he emphasized the importance of soil-slurry conductivities that give patterns at least as good as those from trace metals, that arise from the same electrochemical processes, and that are easier to determine in the field¹⁴. Patterns of conductivity are clarified if the measurements are corrected for H⁺ concentrations that can be derived from pH measurements on the soil slurry. H⁺ concentrations in soil slurries show typical anomalies in profiles across some ore deposits even though the deposits are covered by as much as 10 metres of glacial deposits^{15,16}. Lithium gives a similar anomaly around an ore deposit in Norway probably, according to Bölviken and Logn¹⁷, because Li⁺, like H⁺, has high ionic equivalent conductance and should be one of the principal current carriers in an electrochemical system.

Many other processes besides electrochemical ones influence the formation of ore deposits as well as the zoning of metals around them, but an understanding of electrochemical effects may explain the origin of some ore deposits and provide a means for exploring for them. The work reviewed here has been done mostly by scientists in government and university research organizations, with extensive cooperation from mining companies to which the subject is of great interest and potential value. □



100 years ago

The Jablochhoff light has been introduced by M. Hervé-Mangon into the Conservatoire. It will be fed by a Gramme machine, which the establishment has purchased for its constant use. The light will be placed in the amphitheatre, where M. Hervé-Mangon delivers, twice a week, his own lectures.

The lighting of the Victoria Station of the District Railway by means of the Jablochhoff electric light has been so successful that it has been also applied to the Charing Cross Station, and will shortly be introduced at Earl's Court.

From *Nature* 23, 18 November, 64, 1880.

- Morris, R.C., Thornber, M.R. & Ewers, W.E. *Nature* **288**, 250-252 (1980).
- Dobrin, M.B. *Introduction to Geophysical Prospecting* (McGraw-Hill, New York, 1960).
- Jakosky, J.J. *Exploration Geophysics* (Trijia, Los Angeles, 1950).
- Sato, M. & Mooney, H.M. *Geophysics* **25**, 226 (1960).
- Pourbaix, M. *Thermodynamics of Dilute Aqueous Solutions* (Arnold, London, 1949).
- Latimer, W.M. *The Oxidation States of Elements and their Potentials in Aqueous Solutions*, 2nd edn (Prentice-Hall, New York, 1952).
- Rose, A.W., Hawkes, H.E. & Webb, J.S. *Geochemistry in Mineral Exploration* (Academic, New York, 1979).
- James, H.L. & Sims, P.K. (eds) *Econ. Geol.* **68**, 913 (1973).
- Dorr, J. van N. II *Econ. Geol.* **59**, 1203 (1964).
- Constantinou, G. & Govett, G.J.S. *Trans. Instn. Min. Metall.* **81**, B34 (1972).
- Govett, G.J.S. *J. Geochem. Explor.* **1972** (Instn. Min. Metall., London, 1973).
- Govett, G.J.S. & Whitehead, R.E.S. *Econ. Geol.* **69**, 551 (1974).
- Thornber, M.R. *Chem. Geol.* **15**, 1, 117 (1975).
- Govett, G.J.S. *J. Geochem. Explor.* **1974** (Elsevier, Amsterdam, 1975).
- Govett, G.J.S. *J. Geochem. Explor.* **6**, 359 (1976).
- Govett, G.J.S. & Chork, C.Y. *Prospecting in Glaciated Terrain* (Instn. Min. Metall., London, 1977).
- Bölviken, B. & Logn, Ö. *J. Geochem. Explor.* **1974** (Elsevier, Amsterdam, 1975).

REVIEW ARTICLE

Do waves limit turbulent diffusion in the ocean?

John D. Woods

Institute für Meereskunde an der Universität Kiel, Düsternbrooker Weg 20, D2300 Kiel 1, FRG

Recent theoretical and observational studies of fluctuating motions in the stable interior of the ocean suggest that the rate of turbulent diffusion of scalars may be sensitive to the influence of waves on the energy balance of eddies and fronts. A new flow diagram is used to describe the turbulent cascades of energy and enstrophy in the ocean.

REYNOLDS introduced the idea of separating motion into two parts, steady and fluctuating. In the ocean the gyre circulations can be thought of as the steady part and all the other motions as the fluctuating part. This division is clear in the climatological mean spectrum of oceanic kinetic energy (Fig. 1). The fluctuating part is normally divided into two categories.

(1) Wave-like motions that radiate energy and momentum through the ocean. Their propagation through the ocean can be described kinematically in terms of linear theory, plus resonant wave-wave interaction and dynamical interaction with the environment through which they pass.

(2) Advective-like motions that can only transport momentum and energy by advection of water particles from one place to another. The fluctuating motions in this category are often sufficiently energetic to require nonlinear description of their development with time. They can be further subdivided into (i) those motions that do not overturn isentropic surfaces (notably quasi-geostrophic eddies, and fronts); (ii) those that do overturn isentropic surfaces (billows, Langmuir cells, and the motion within patches of three-dimensional turbulence).

Many fluid dynamicists prefer to restrict the term 'turbulence' to the category (2ii). In this limited definition oceanic turbulence occurs only at the boundary layers, in deep-convection and in thin, transient blini in the interior. Oceanographers, however, talk about lateral diffusion in the interior of the ocean, occurring at rates ($\sim 1,000 \text{ m}^2 \text{ s}^{-1}$) that cannot possibly be explained by the restricted picture of the fluid dynamicist. Oceanographers accept that the fluctuating motions in category (2i), those that do not overturn isentropic surfaces, demand statistical treatment in the spirit of turbulence theory. They too must be included in any description of 'turbulence in the ocean'. Recently, there has been increasing discussion about the need to include wave-like as well as advective-like fluctuations in any complete model of turbulence in the ocean designed to predict the rates of scalar diffusion. The basis for this is that the rate of turbulent diffusion depends on the kinetic energy density of the advective-like fluctuations which is limited by leakage of energy to waves that radiate the surplus energy away and dissipate it.

This article summarizes the main features of the conceptual model of ocean turbulence that emerges if waves play a significant part. The problem can be simplified by ignoring the well documented regional variation of turbulent kinetic energy. The method, a common one in turbulence theory, considers the statistics of fluid-dynamically important properties such as momentum and vertical component of vorticity, and in particular describes how their variances (kinetic energy and enstrophy, respectively) are transferred from one place to another and from one scale to another. The transfer from one place to another is achieved by advection (by mean circulation) and by radiation (of energy but not enstrophy) by waves. The transfer from one scale to another is called a cascade, which is a transfer

from one position to another in the wavenumber spectrum. We can also think of it as a flux of variance (energy, enstrophy, and so on) from one place to another in Fourier space. An integrated picture of turbulence in the ocean is developed by considering the budgets of kinetic energy and enstrophy, taking account of their fluxes through both physical and Fourier space by advection, radiation and cascading. To achieve this conservation laws are used which are like the laws of thermodynamics, but concerned with kinetic energy and enstrophy of the fluctuations of oceanic motion. There are important relations between the conservation laws of thermodynamics and turbulence theory. The most important law of ocean turbulence is that potential vorticity is conserved in adiabatic motion, that is, those that do not involve overturning and subsequent mixing of isentropic surfaces. Oceanographers have long relied on this law in developing theories of the gyre scale mean circulation; it is the basis of the Sverdrup balance. The law is also valid for fluctuating motions with scales ranging down to the order of 1 m in the interior of the ocean. This is because the vertical density gradient in the ocean interior is so large that the rate of kinetic energy input (of the order of mW m^{-2}) is only sufficient to sustain intermittent overturning of isentropic surfaces on scales of 1 m or less. Between these rare and short-lived overturning events there is no three-dimensional turbulence; so, at any instant the interior of the ocean is in laminar flow almost everywhere. The law of potential vorticity conservation can be applied across the spectrum from the scale of the ocean basin down to the scale of overturning (of the order of 1 m). The climatological spectrum in Fig. 1 shows that this covers almost the whole of the range now under discussion. Conservation of potential vorticity is an approximation but the largest source of error—intermittent overturning and mixing of isentropic surfaces in the interior—is negligible for the present discussion. There is a striking similarity between this law and the law of conservation of relative vorticity in a strictly two-dimensional flow and many of the results of two-dimensional turbulence theory can be applied to oceanic turbulence. In discussing the statistics of the fluctuating oceanic flow we shall therefore use 'enstrophy' to describe 'variance of potential vorticity'.

The gyres

In this article the main features of the new conceptual model of oceanic cascade physics are summarized using a new flow diagram presented in Fig. 2. On the left lies the largest scale of motion, the gyres of the upper ocean (A), driven by the wind through the agency of Ekman pumping (a). The gyres exhibit seasonal variation in response to the wind, but we can think of their streamlines as being steady compared with the fluctuating motions associated with the smaller-scale eddies, fronts, waves, and so on. The gyre circulation represents the mean flow, in

terms of Reynolds's division of a turbulent flow. Between adjacent gyres there are relatively sharp permanent fronts, characterized by maxima in both thermoclinicity and baroclinicity. The gyre streamlines pack closer in these frontal jets and in western boundary currents (for example, the Gulf Stream). Water particles take a decade or so to circulate around an upper ocean gyre. The size, location and number of gyres remains controversial², so we are discussing the physics of the fluctuations without knowing the mean flow. This difficulty arises largely from uncertainty over the influence of the fluctuations on the mean motion; a first-order problem in both diagnostic and prognostic models of oceanic circulation.

The eddies

Hydrodynamic instability (b) of the gyres^{3,4} and unsteady external forcing effects (wind, surface heat and moisture transfer) generate eddies (B) whose horizontal dimensions are of the order of 100 km and circulation speed of the order of 10 km

day⁻¹. These eddies have a Rossby number ($R_0 = U/Lf$) of the order of 0.01 and are, therefore, quasi-geostrophic. Field observations⁵ during the 1970s have shown that such eddies occur widely (believed throughout the ocean) albeit with considerable regional variation (being strongest around the baroclinicity maxima of permanent fronts and western boundary currents). They are readily detected in oceanographic sections by the order of 100 m vertical displacement of isopycnals (and isotherms). The energy density in the spectral peak associated with the eddies (Fig. 1) exceeds that of the gyres by a factor of nearly 100; they are the energy-containing eddies of oceanic turbulence⁶. As detached eddies form from meanders, they interact with the gyre scale circulation in the upper ocean, influencing the pattern of mean streamlines⁷. There is some evidence⁸ that, like their atmospheric counterparts⁹, eddy stress profoundly influences (c) the general circulation as though by negative viscosity. On the other hand, the eddies spread passive scalar anomalies along density surfaces at a rate equivalent to a (positive) diffusivity of $\sim 70 \text{ km}^2 \text{ day}^{-1}$ (refs 6, 10).

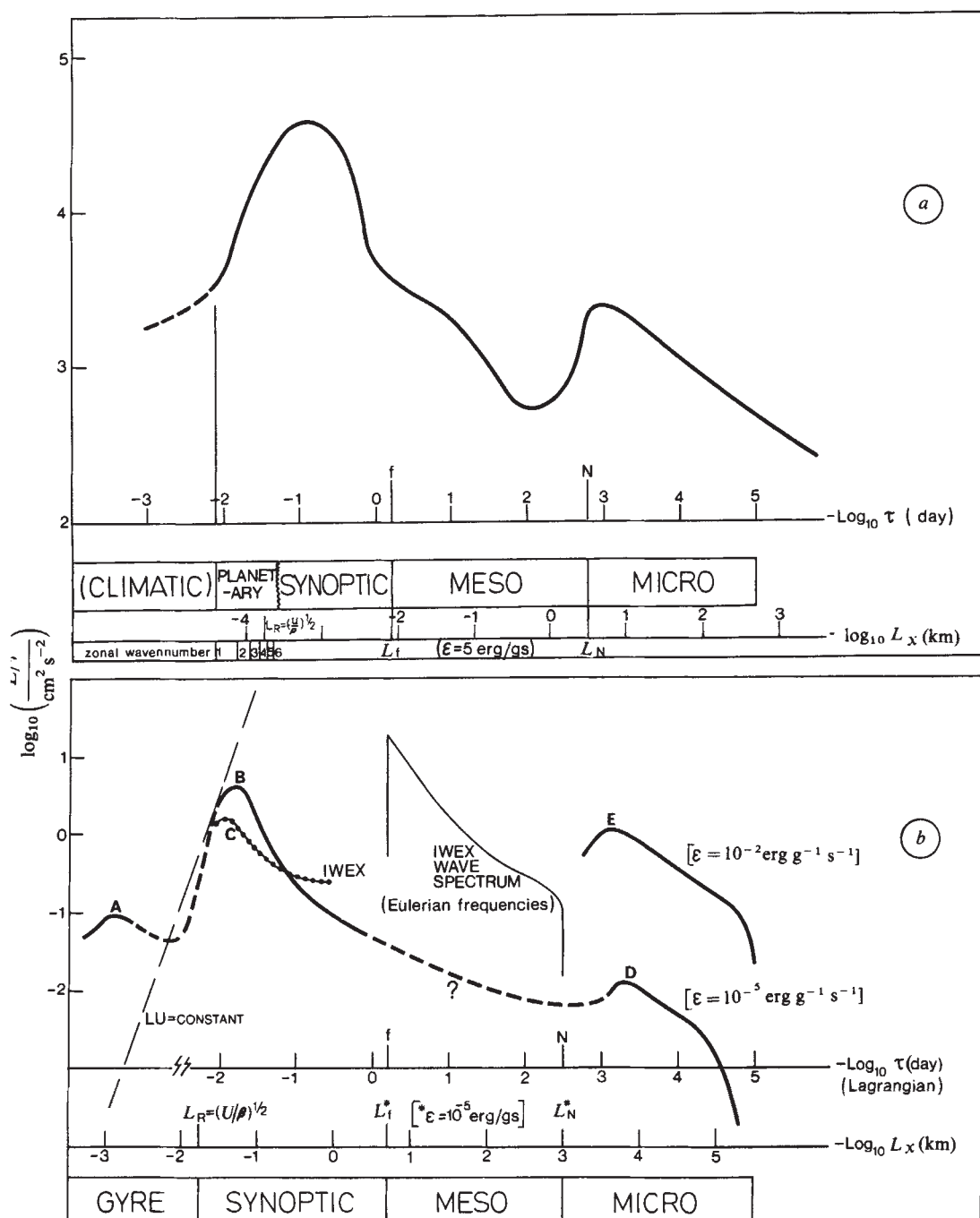


Fig. 1 A comparison of the climatological spectra for: *a*, turbulent kinetic energy in the atmosphere; *b*, turbulent kinetic energy and internal wave energy in the ocean based on recent observations. The space and time scales are in both cases approximate, being based on a mixture of eulerian and lagrangian estimates. (Figure reproduced from ref. 38.)

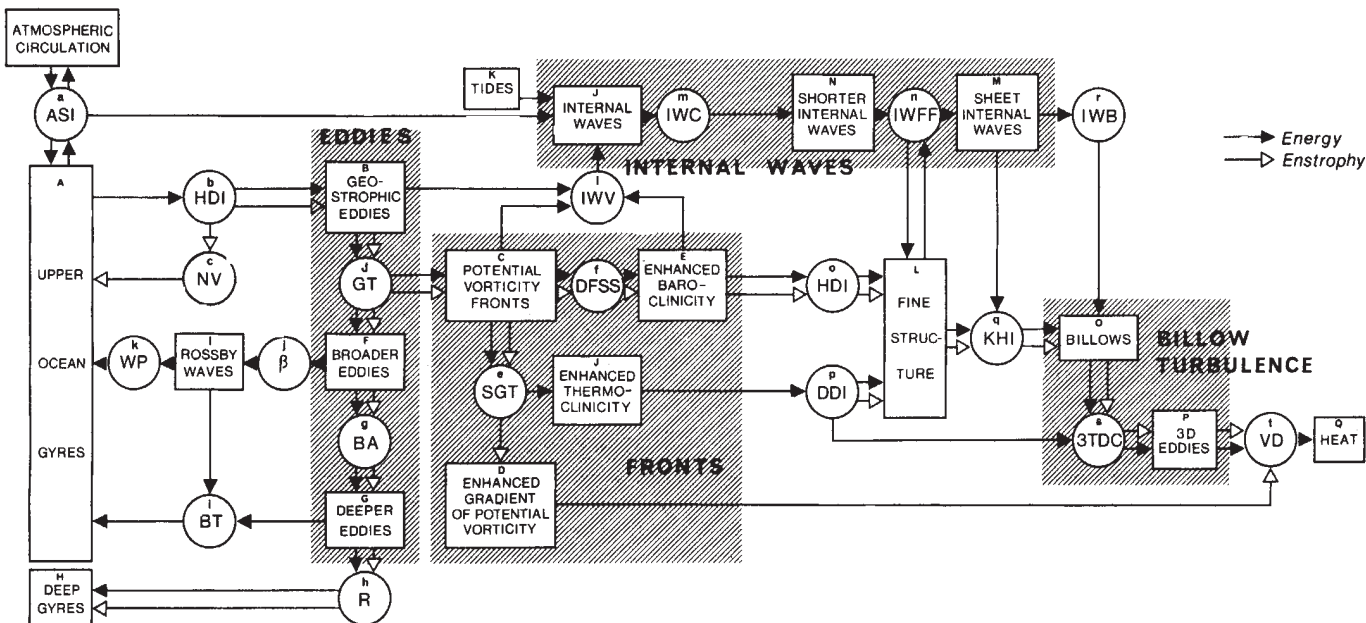


Fig. 2 The new flow diagram for cascades of energy and enstrophy in the ocean. Letters on each box are indexed in the text. The following abbreviations have been used in labelling the circled processes: (a) ASI, air-sea interaction; (b) HDI, hydrodynamic instability (baroclinic and barotropic); (c) NV, negative viscosity, associated with baroclinic instability; (d) GT, geostrophic turbulence, which initiates the process of kinematic frontogenesis; (e) SGT, semi-geostrophic turbulence; (f) DFSS, dynamic frontogenesis at the (baroclinic) sea surface; (g) BA, baroclinic adjustment; (h) R, rectification of deep eddy enstrophy to give deep gyres; (i) BT, interaction with bottom topography; (j) β , interaction between eddies and Rossby waves; (k) WP, westward propagation of Rossby wave energy; (l) IWV, internal wave viscosity; (m) IWC, internal wave cascade of energy to higher wavenumbers; (n) IW-FF, interaction between internal waves and fine structure—internal wave strain produces fine structure, and fine structure permits very short trapped 'sheet' waves; (o) HDI, hydrodynamic instability of mesoscale fronts; (p) DDI, double-diffusive instability, leading to convection; (q) KHI, Kelvin-Helmholtz instability; (r) IWB, internal wave breaking (other than KHI); (s) 3DTC, three-dimensional turbulence cascade; (t) VD, dissipation of turbulent kinetic energy and enstrophy by molecular viscosity.

Eddy decay

Observations of the streamline patterns associated with the eddies show a slow evolution with a lagrangian correlation time scale of ~ 50 days (ref. 6); so, although they are the oceanic equivalent of the synoptic scale storms of the atmosphere, the eddies decay very much more slowly. (Rings, as distinct from eddies, decay an order of magnitude more slowly). One of the principal challenges in the study of oceanic turbulence is to develop a theory for this observed rate of decay. To decay, the eddies must lose both energy and enstrophy. The observed lifetimes indicate that this must occur at rates given, to order of magnitude, by the following example:

Eddy width

$$L = 100 \text{ km}$$

Rotation speed

$$U = 10 \text{ km day}^{-1}$$

Energy density

$$E = \frac{U^2}{2} = 50 \text{ km}^2 \text{ day}^{-2} \text{ per unit mass}$$

Enstrophy density

$$\Omega' = \frac{U^2}{2L^2} = 5 \times 10^{-3} \text{ day}^{-2} \text{ per unit mass}$$

Observed lifetime

$$T = 100 \text{ days}$$

Deduced dissipation rates:

$$(1) \text{ Energy } \quad \epsilon = 0.5 \text{ km}^2 \text{ day}^{-3}$$

$$(2) \text{ Enstrophy } \quad \eta = 5 \times 10^{-5} \text{ day}^{-3}$$

Ultimately the energy and enstrophy are dissipated by molecular viscosity ($\nu = 10^{-2} \text{ cm}^2 \text{ s}^{-1}$ for seawater). Neglecting, for the moment, the complex processes shown in Fig. 2, we can estimate dissipation scales for the above example.

Dissipation scales:

$$(1) \text{ Energy } \quad l_e = (\nu^3 \epsilon^{-1})^{1/4} = 1 \text{ cm}$$

$$(2) \text{ Enstrophy } \quad l_n = (\nu^{1/2} \eta^{-1/6}) = 1 \text{ m}$$

In this primitive model, the eddy energy and enstrophy have to cascade through seven and five decades respectively of Fourier space before molecular viscosity can dissipate them, a requirement not significantly changed by the new model.

The enstrophy cascade

The basic mechanism of the enstrophy cascade has been recognized for well over a decade, being closely related to the theoretical idealization of inviscid two-dimensional turbulence, in which particles retain their relative vorticity because there is no vortex stretching. Although the eddy motion in the interior of the ocean is not strictly two-dimensional it is—to a much better approximation—adiabatic, with the result that particles effectively retain their potential vorticity as they are swept around the complicated, fluctuating eddy streamline patterns. Below the wind-mixed layer and below the top few metres of the ocean, where the Sun warms the water diabatically, the most serious cause of error in the assumption of potential vorticity conservation arises from diabatic mixing by motion that overturns isentropic surfaces. Identifying this with billow turbulence we can estimate the magnitude of the error. The time scale for potential vorticity change due to diabatic mixing at rate K , in a region where the curvature of the vertical profile of N has a scale depth L_N , is given by $T \approx L_N^2/K$. Taking $L_N = 0.1 \text{ km}$ for the oceanic thermocline, and estimating K from the above example

$$K = \epsilon N^{-2} = 5 \times 10^{-7} \text{ km}^2 \text{ day}^{-1}$$

where $N = 1,000 \text{ day}^{-1}$, we calculate that $T \sim 2 \times 10^5$ days, which is so much longer than the eddy lifetime that we can neglect the effect of diabatic mixing. (The same calculation also justifies the assumption of potential vorticity conservation in gyre dynamics, for which the circulation time is of the order of 10 yr.)

Fronts

The enstrophy cascade is, therefore, assumed to be adiabatic, and to proceed by lateral redistribution of potential vorticity¹². As Batchelor¹¹ showed in the simpler case of two-dimensional turbulence, this leads to kinematic frontogenesis (d). A multiplicity of ever-extending, ever-thinning fronts (C) develop in the eddy contours of constant potential vorticity. This advective redistribution of eddy enstrophy has important consequences. First, the stretching of contours of constant potential vorticity into long, thin fronts introduces a local anisotropy involving sharp gradient of vorticity (D) and correspondingly large Rossby number in the cross-front direction. The resulting flow field is described by the semi-geostrophic equations used in studies of mesoscale fronts^{13,14}. We therefore call this motion occupying the waveband 10 km to 1 m 'semi-geostrophic turbulence'. Most of the kinetic energy in each front is in its horizontal, quasi-geostrophic jet; however, the weak ageostrophic cross-front motion associated with the continuous thinning of the front is not horizontal, but rather it runs up and down the density surfaces inclined by the eddy baroclinicity. (Roughly speaking, the axes of the fronts lie tangentially to the adjacent eddies, like the bands of a spiral nebula.) At the sea surface, where density surfaces outcrop, this ageostrophic motion redistributes the strong near-surface thermoclinicity (due to solar heating) to give the characteristic modulation of sea surface temperature seen in



Fig. 3 An IR image of the central Mediterranean Sea from NOAA 5 showing eddies sharply outlined by sea-surface temperature gradients at mesoscale fronts. (Photograph reproduced by permission of Dr P. Baylis, University of Dundee.)

satellite IR images (Fig. 3). At the same time, the kinematic frontogenesis (d, e) forced by the eddy motion interacts with the upper boundary (f) causing the frontal baroclinicity to peak strongly in the top few decametres, but to flatten below¹⁴ (E), so that the kinetic energy of the semi-geostrophic turbulence becomes concentrated near the sea surface.

Eddy expansion

As the eddy vorticity becomes teased out into fronts that extend beyond the original dimensions of the eddy, the mean eddy circulation also expands (F), carrying with it a small portion of the original eddy vorticity¹². This lateral expansion is the oceanic realization of the reverse energy cascade in two-dimen-

sional turbulence theory. The broadening introduces baroclinic adjustment (g) with the result that eddy potential energy is transferred to eddy kinetic energy; the eddy rotates faster. At the same time, due to potential vorticity conservation, the vertical distribution of eddy energy tends to become more uniform; the eddy becomes more barotropic (G). This drives eddy enstrophy down into the deep ocean, where it can be rectified (h) to force deep gyre circulations¹⁵ (H). As the eddies deepen they also begin to feel the bottom topography (i) producing the observed spatial inhomogeneity⁵ and providing the link between the mean circulation of the upper ocean gyres and the much deeper bottom topography. Hence the steering of the Gulf Stream by bottom topography is a consequence of frontogenesis in the Gulf Stream eddies.

Rossby waves

As an eddy broadens it soon approaches the limit beyond which the North-South difference in planetary vorticity across the eddy exceeds the mean eddy (relative) vorticity

$$\beta L > U/L$$

For the eddy in our example, this occurs when $L > 200$ km. At this scale the eddy motion becomes influenced by the rotating Earth's sphericity (j) and energy begins to radiate as Rossby waves¹⁶ (I), which propagate westwards. Streamline patterns derived from the MODE data exhibit this characteristic westward phase drift⁵. Compared with the full spectrum from millimetres to megametres, the eddies expand very little before Rossby wave radiation takes hold. The proximity of the scales at which the eddies first receive energy from the gyres and subsequently are converted to Rossby waves explains the narrowness of the observed spectral valley (not quite a 'gap') between the spectral peaks of the eddies and the gyres (Fig. 2). The Rossby waves powered by eddy expansion radiate energy to the distant ocean (possibly after scattering by bottom topography (i)), allowing mean circulation to be driven from a distance. The effect may be intensified at a western boundary, where energy tends to be trapped (k), accelerating the western boundary current and so contributing to the associated gyre circulation^{15,16}.

Internal waves

In classical inviscid, two-dimensional turbulence theory^{17,18} all eddy energy is transferred to larger scale and eventually to the gyres, but in the ocean much of the energy cascades to smaller scale to be dissipated locally by molecular viscosity. The mechanism for this energy cascade cannot be turbulent frontogenesis for, although it efficiently cascades enstrophy to small scale, the associated energy cascade rate, ϵ_t , is reduced in proportion to the square of the dissipation scale,

$$\epsilon_t = \eta_t l^2$$

For our example ($\eta_t = 5 \times 10^{-5} \text{ day}^{-3}$, $\epsilon = 0.5 \text{ km}^2 \text{ day}^{-1}$, $l_e = 10^{-5} \text{ km}$)

$$\epsilon_t = 5 \times 10^{-15} \text{ km}^2 \text{ day}^{-1} = 10^{-10} \epsilon$$

Thus, frontogenesis cannot dissipate much of the eddy energy. On the other hand, observations of ocean microstructure^{20,21} are broadly consistent with the mean energy dissipation rate of our example ($\epsilon = 0.5 \text{ km}^2 \text{ day}^{-1}$). How then does energy cascade from the eddies to small scale? One solution is that internal waves (J) extract energy from the eddies (and from the fronts) through the Reynolds stress they exert on the shear in the eddy/front flow²². The magnitude of the corresponding 'internal wave viscosity' (I) has recently caused controversy^{23,24}, earlier estimates of $0.1 \text{ km}^2 \text{ day}^{-1}$ being later revised downwards by a

factor of 10–100. The higher values were estimated on the basis of dynamical theory; the lower values are based on field observations. Nevertheless, the concept of internal wave viscosity is generally accepted, only the magnitude is in doubt. Taking an intermediate value of $K_w = 0.005 \text{ km}^2 \text{ day}^{-1}$, we estimate the rate of energy extraction from eddies with a mean vertical shear of 10 day^{-1} to be

$$\varepsilon_w = K_w (dU/dz)^2 = 0.5 \text{ km}^2 \text{ day}^{-3}$$

This is precisely the value of ε in our example, so we tentatively conclude that internal wave viscosity provides an important mechanism for extracting energy from the eddies.

Once the energy enters the internal wave spectrum it is believed to cascade efficiently by wave–wave interaction to the smallest scales of wave motion. This concept of an internal wave energy cascade is based on a combination of dynamical theory and empirical knowledge acquired during the past decade. In particular, observations have revealed a remarkable homogeneity of spectral form and level in the interior of the ocean²⁵, with significant modifications in the upper ocean²⁴. This is interpreted in terms of the universal saturation of the internal wave field²³. Whenever additional energy is added at any scale (for example, by interaction with an eddy) the surplus is rapidly cascaded through the spectrum by wave–wave interaction (m) until it reaches the smallest scales, where the waves break, transferring the energy to three-dimensional turbulence²⁴. Orlanski²⁶ has suggested that the shedding of surplus energy may occur even more efficiently by the enhanced large-scale wave strain causing small waves to break directly without a cascade through the whole wave spectrum. Either way, the large (1–10 km) internal waves that extract energy from the eddies/fronTS pass it to small scales (~5 m) so efficiently that the energy is not propagated far from the point of injection before it is lost to breaking²⁷. Energy removed from eddies by internal wave viscosity is dissipated locally, probably within some tens of kilometres. Note that the internal wave spectrum is not kept saturated by the energy loss from the eddies; there are other important sources, including air–sea interaction (a) and tides (K). Nor need the energy extraction from eddies come exclusively from internal wave viscosity: a significant part may come from the interaction of the eddies with bottom topography (not shown in Fig. 2). Finally the local character of energy dissipation by internal waves suggests that observed spatial inhomogeneity in the internal wave field, especially in the upper ocean, may reflect recent interactions with eddies and, especially in the upper ocean, with fronts²⁸ (E).

Fine structure

Turning back to Fig. 1, we now arrive at the fine structure box (L). We have no theory for ocean fine structure just observations interpreted speculatively on the basis of analogy with laboratory experiments or otherwise. Some of the observed fine structure in density stratification can be explained by the transient straining (n) of the mean density field by internal waves²⁹. The thermohaline characteristics of the fine structure suggests that water masses of different characteristics have intruded into one another in layers with vertical scales in the range 1–100 m and horizontal scales of up to several tens of kilometres²⁴. Such structure is particularly sharply defined, with thin sheets (vertical scale ~10 cm) separating the layers, near fronts where the mean cross-front thermoclinicity is enhanced by kinematic frontogenesis²⁸ (J). A correlation between intrusions and meanders on fronts has suggested that the dynamical instability (o) of mesoscale fronts may provide an important mechanism for fine structure generation³⁰. Theoretical studies of frontal instability showing complex vertical structure in the growing waves support this conjecture (M. K. MacVean, personal communication). An alternative proposal is that intrusions are driven by double diffusive convection (p), which is also most

effective in the enhanced thermoclinicity at fronts³¹. Most observations of fine structure comprise single profiles or sections, which do not provide the information needed to discriminate between these three main mechanisms: (1) internal wave straining (n); (2) dynamical intrusion associated with frontal baroclinicity (o); (3) convective intrusion driven by the double diffusive release of latent static instability associated with frontal thermoclinicity (p). New techniques yielding time series of three-dimensional surveys of fine structure may help to overcome these limitations³².

The net influence of fine structure on the overall cascades of energy and enstrophy in the ocean remains uncertain, but its presence significantly modifies the details of the important transition from the cascades of frontogenesis and wave–wave interaction, in which isentropic surfaces are not overturned, to

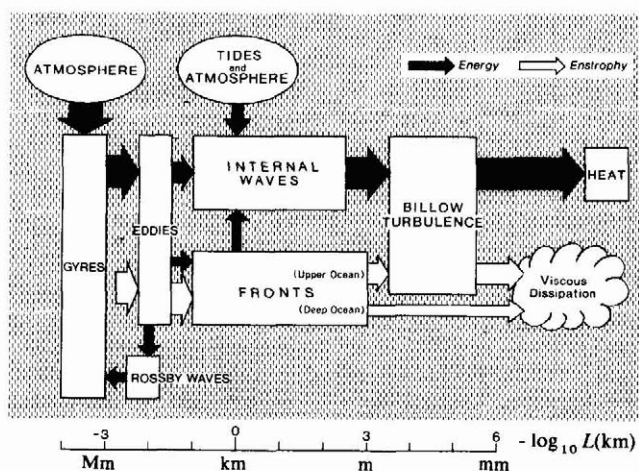


Fig. 4 A simplified cascade diagram showing the principal paths by which oceanic eddy energy and enstrophy flow through the spectrum. The scale provides an indication of the approximate positions of the various classes of motion in the overall spectrum megametres to millimetres. The breadths of the cascade arrows are intended to suggest a possible distribution between competing paths.

the cascade of three-dimensional quasi-isotropic turbulence, in which they are. I envisage fine structure as the bridge between these two classes of cascade; not a single-span bridge—more like a complicated tropical bamboo construction. We may be able to develop theories for the statistical properties of ocean fine structure on the basis of some thermodynamic-like principle of optimum energy and enstrophy transfer between the three cascades. But for the present we must merely note elements of the fine structure mechanism. First, the internal wave community has drawn attention to the importance of fine structure in determining the high wavenumber cutoff of the internal wave spectrum³³. Flow visualization observations have shown that the shortest internal waves—'sheet waves' (M)—are trapped on the thin sheets of enhanced static stability. It is not clear how they are generated; presumably by interaction (n_2) between the fine structure and the shortest free internal waves. The trapped sheet waves are important because they and not the free waves are observed to break³⁴. Breaking has been observed to result from Kelvin–Helmholtz instability (q), which occurs when the local Richardson number is <0.25 .

$$Ri = \left| N / \frac{\partial U}{\partial z} \right|^2, \quad \left(\text{where } \frac{\partial U}{\partial z} \text{ is the shear} \right)$$

Kelvin–Helmholtz instability occurs infrequently when a sheet wave provides additional fluctuating shear (due to the orbital motion of the water particles) to supplement the more steady

shear of the fine structure currents³⁴. Wave breaking may also occur independently of shear instability (r)³⁵, but this has not been observed in the ocean.

Billow turbulence

Kelvin-Helmholtz instability causes density (and isentropic) surfaces to overturn giving billows (O). Observations show that billows occur in rows on a sheet making blini several metres across in otherwise laminar flow³⁶ (defined here as being flow devoid of motion that overturns density surfaces). Empirical estimates of the instantaneous volume fraction occupied by billows in the upper ocean range from a few (~ 5) per cent in calm weather to an order of magnitude greater in stormy weather³⁷. There is also considerable regional variability, with the highest concentrations found at intense currents and near fronts^{20,21,38}. This intermittency must be allowed for as we follow the cascade of energy from the non-overturning motions (internal waves, fine structure currents) into the overturning motions of billow turbulence.

In making this transition we abandon the assumption of potential vorticity conservation and no longer consider the enstrophy cascade which becomes lost in the explosive generation of enstrophy by billow turbulence. The dual cascade of the mesoscale becomes fused into one as we pass into the microscale, thanks to the three-dimensional process of vortex stretching, which increases enstrophy at a rate that allows energy to cascade with no flux divergence until it becomes scattered into heat by molecular viscosity. The billows, the energy-containing eddies of three-dimensional turbulence in the interior of the ocean, are observed to be only a few decimetres high. They have Reynolds numbers of a few hundreds, barely sufficient to permit a modest inertial cascade of isotropic turbulence inside each billow (r). The average energy cascade rate, $\bar{\epsilon}_B$, inside such tiny cylinders of three-dimensional billow turbulence is related to the mean energy flux $\bar{\epsilon}_w$ through the internal wave cascade (neglecting contributions from sources other than the eddies) by the mean billow intermittency factor $\bar{I}$: that is

$$\bar{I}\bar{\epsilon}_B \cong \bar{\epsilon}_w$$

For $\bar{I} = 0.05$ and $\bar{\epsilon}_w = 0.5 \text{ km}^2 \text{ day}^{-3}$, $\bar{\epsilon}_B = 10 \text{ km}^2 \text{ day}^{-3}$, which is typical of direct estimates from velocity microstructure measurements^{20,21}. For this value of the billow turbulence energy cascade rate, the Kolmogorov microscale of molecule dissipation is

$$l_e = (\nu^3 \bar{\epsilon}_B^{-1})^{1/4} = 3 \text{ mm}$$

And the Ozmidov scale of largest overturning (taking $N = 1,000 \text{ day}^{-1}$) is

$$l_N = (\epsilon_B / N^3)^{1/2} = 10 \text{ cm}$$

in good agreement with direct observation.

Conclusion

The conversion of turbulent kinetic energy to heat (Q) and the associated molecular dissipation of enstrophy (t) complete the cascades begun with the creation of geostrophic eddies by instability of the mean gyre circulation. The cascades span 10 decades of Fourier space, a waveband one hundred million times broader than that of the most advanced prognostic model of ocean circulation. I have discussed the physics of processes that influence the mean circulation but which can never be resolved in such models. These unresolved processes must be treated statistically in the model equations—their effect on the large-scale circulation and distributions must be parameterized. Current practice relies on eddy transport coefficients invented by Boussinesq in the nineteenth century. Systematic investigation of the physics of ocean turbulence began in the 1960s and has progressed rapidly during the 1970s, with major advances in our understanding of eddies, waves, fronts, fine structure and billows. The new cascade flow diagram is not claimed to be correct, but it seems to offer the simplest conceptual model consistent with contemporary theoretical ideas and with the admittedly still primitive estimates of cascade rates. I anticipate and invite modifications. For example, the role of the ocean surface in modifying the cascade physics has scarcely been mentioned and the role of the seasonal change not at all. The interior of the ocean has been emphasized rather than the boundary layers, although these are likely to be important. Finally, according to the model, energy is extracted from the eddies by both internal waves and Rossby waves, but the latter only after the eddies have expanded by a factor of about two. If the internal wave viscosity is higher in some regions, a larger proportion of eddy energy will cascade to billow turbulence leaving less to feed Rossby waves, and vice versa. Does this imply less eddy expansion in regions of greater wave viscosity? If so, then there will also be less eddy deepening in these regions, and therefore weaker interaction with bottom topography and more feeble powering of deep gyres. But the wave viscosity depends on the internal wave saturation level and therefore on the mean static stability (and perhaps on the fine structure intensity). This suggests that the cascade physics described by the proposed new flow diagram should be inherently stable when applied to oceanic circulation models.

This article is based on a lecture presented at the IAPSO General Assembly, Canberra, December 1979 and uses material presented at the Second International Symposium on Turbulence in the Ocean (Liege, May 1979) co-sponsored by IAPSO and SCOR with support from UNESCO, IOC and IDOE. (Contributed papers are reported in ref. 39 and invited papers in ref. 40.) I thank Konstantin Fedorov, Terry Joyce, Sergei Kitaigorodskii, Dirk Olbers, Peter Rhines and Stewart Turner for helpful comments on an earlier version of the article.

1. IOC Workshop Rep. No. 21 (UNESCO, 1979).
2. Worthington, V. *On the North Atlantic Circulation* (Johns Hopkins, Baltimore, 1976).
3. Gill, A. E., Green, J. S. A. & Simmons, A. *Deep-Sea Res.* **21**, 499–528 (1974).
4. Rhines, P. B. in *The Sea* Vol. 6, 189–318 (Wiley-Interscience, London, 1979).
5. The MODE group *Deep-Sea Res.* **25**, 859–910 (1978).
6. Freeland, H. J., Rhines, P. B. & Rossby, H. T. *J. mar. Res.* **33**, 383–404 (1975).
7. Holland, W. R. in *The Sea* Vol. 6, 3–46 (Wiley-Interscience, London, 1979).
8. Webster, F. *Deep-Sea Res. Suppl.* **16**, 357–368 (1969).
9. Green, J. S. A. *Q. Jl R. met. Soc.* **96**, 157–185 (1970).
10. Veronis, G. in *The Sea* Vol. 6, 169–188 (Wiley-Interscience, London, 1979).
11. Batchelor, G. K. *Phys. Fluids Suppl.* **12**, 233–238 (1969).
12. Rhines, P. B. A. *Rev. Fluid Mech.* **11**, 401–441 (1979).
13. Hoskins, B. & Bretherton, F. P. *J. atmos. Sci.* **29**, 11–37 (1972).
14. MacVean, M. K. & Woods, J. D. *Q. Jl R. met. Soc.* **106**, 293–311 (1980).
15. Holland, W. & Rhines, P. B. *J. phys. Oceanogr.* (submitted); *Dyn. Atmos. Oceans* **3**, 289–325 (1979).
16. Rhines, P. B. *J. Fluid Mech.* **69**, 417–443 (1975).
17. Taylor, G. I. *Phil. Trans. R. Soc. A* **240**, 1–26 (1915).
18. Fjortoft, A. *Tellus* **5**, 225–230 (1953).
19. Lilly, D. in *Dynamic Meteorology* (Reidel, Dordrecht, 1974).
20. Gibson, C., Vega, L. A. & Williams, R. B. in *Advances in Geophysics* **18A**, 353–370 (Academic, London, 1974).
21. Monin, A. S. & Ozmidov, R. V. in *Physics of the Ocean* **1** (Woods, Moscow, 1978).
22. Müller, P. & Olbers, D. *J. geophys. Res.* **80**, 3848–3860 (1975).
23. Garrett, C. J. & Munk, W. H. A. *Rev. Fluid Mech.* **11**, 339–69 (1979).
24. Gregg, M. & Briscoe, M. G. *Rev. Geophys. Space Phys.* **17**, 1524–1528 (1979).
25. Müller, P., Olbers, D. & Willebrand, J. *J. geophys. Res.* **83**, 479–500 (1978).
26. Orlandi, I. in *Marine Turbulence* 65–100 (Elsevier, Amsterdam, 1980).
27. Olbers, D. in *Turbulence in the Ocean* (Springer, Berlin, in the press).
28. Woods, J. D. in *Physics of Fronts in the Ocean* (in preparation).
29. de Saubies, Y. *Proc. 14th Liège Ocean Hydrodynamics Colloq.* (Elsevier, Amsterdam, 1980).
30. Woods, J. D., Wiley, R. L. & Briscoe, M. G. in *A Voyage of Discovery*, 253–275 (Pergamon, Oxford, 1978).
31. Turner, J. S. in *Turbulence in the Ocean* (Springer, Berlin, in the press).
32. Woods, J. D. & Minnett, P. J. *Deep-Sea Res.* **26A**, 85–96 (1979).
33. Garrett, C. J. R. *Dyn. Atmos. Oceans* **3**, 239–266 (1979).
34. Woods, J. D. *J. Fluid Mech.* **32**, 791–800 (1968).
35. Thorpe, S. A. *J. geophys. Res.* **80**, 328–338 (1975).
36. Woods, J. D. in *Turbulence in the Ocean* (Springer, Berlin, in the press).
37. Nasmyth, P. *ICES Symp. Rep. No. 6* (ICES, 1969).
38. *Proc. IOC-SCOR Study Conf. on Models of Oceanic Circulation and Climate* (WMO, Geneva, 1977).
39. Nihoul, J. C. J. (ed.) *Marine Turbulence* (Elsevier, Amsterdam, 1980).
40. Woods, J. D. (ed.) *Turbulence in the Ocean* (Springer, Berlin, in the press).

ARTICLES

Latitudinal beaming of planetary radio emissions

Dyfrig Jones*

Space Science Department, ESA, Estec, Noordwijk, The Netherlands

The Voyager missions have revealed that jovian kilometric radiation (emanating from Jupiter's magnetosphere) is beamed away from the zenomagnetic equator. Results from GEOS 1 show that terrestrial non-thermal continuum or myriametric radiation is similarly beamed away from the geomagnetic equator. A mode-coupling mechanism is proposed such that measurements of the direction of propagation of the escaping O-mode radiations may allow the construction of the first high-resolution maps of the shapes of the equatorial plasmopause and Io torus.

THE two Voyager spacecraft, which flew by Jupiter in March and July 1979, have shown that Jupiter's magnetosphere has many similarities to the Earth's as far as radio-wave observations are concerned¹⁻⁴. Close parallels had already been drawn between jovian decametric emission, which has been observed from Earth for over two decades, and terrestrial kilometric radiation which has only recently been seen by Earth-orbiting spacecraft⁵. (Decametric and kilometric refer to the free-space wavelengths at which the radiations are predominantly observed.)

One wave component observed by the Voyager spacecraft—jovian kilometric radiation (JKR)—had seemed to have no counterpart within the Earth's magnetosphere. It is suggested here, however, that JKR is the analogue of the Earth's non-thermal continuum radiation^{6,7}, the latter being termed terrestrial myriametric radiation (TMR). Both radiations seem to have their source in the intense electrostatic waves observed to be closely confined to the magnetic equators of both planets^{8,9} in regions of large plasma density gradients—the Earth's plasmopause and the outer edge of the Io plasma torus.

Terrestrial myriametric radiation

A particularly clear example of TMR was obtained for ~80 min on 4 March (day 63) 1978 by the GEOS 1 S-300 wave experiment which incorporates high resolution on-board correlators^{10,11} and sweep frequency analysers¹² (SFAs). The spectrogram shown in Fig. 1 depicts signals from the 40 m tip-to-tip electric dipole as analysed by the SFAs. The start time is 02.50 UT and the end time is ~04.00 UT. During these observations, the satellite moved from ~5.4 R_E to ~4.2 R_E , where R_E is an Earth radius (6,370 km) and from a magnetic latitude of ~16°N to ~8°N. TMR is the striated emission seen in two frequency bands. Other natural emissions occur in the vicinity of the electron plasma frequency ($f_{pe} = 9 \cdot 10^{-3} N_e^{1/2}$ kHz; N_e is the electron density per m^3) and at the f_q 'resonance' respectively. The increase in plasma frequency occurs because the spacecraft is moving towards the Earth and encountering increasing plasma density as the plasmopause is approached. Weak noise at lower frequencies is probably at $3f_{ce}/2$ (ref. 8), where f_{ce} is the electron gyrofrequency ($f_{ce} = 1.8 \cdot 10^3 B_0$ kHz, B_0 is the magnetic field strength) which must also increase as GEOS 1 moves inwards. Clearly the low-frequency cut-off of TMR cannot be depended on to yield unambiguously the local plasma density, as has been suggested previously⁷.

The TMR increases in strength as the spacecraft approaches the plasmopause where intense radiation occurs in the vicinity of f_{pe} (where f_{pe} is the electron plasma frequency). The striations in the TMR are due to a beat between the satellite spin frequency ~0.2 Hz and the frequency ~0.04 Hz with which the SFAs are sweeping through the frequency range 0–75 kHz. The nulls in

the pattern occur when the dipole antenna used for receiving is aligned parallel to the direction of the emission's propagation, the antenna spin plane being approximately parallel to the equatorial plane. The fact that the pattern is so well-defined for so long is an indication that the radiation is arriving from preferred directions as GEOS 1 moves in towards the plasmopause. Such patterns have been used for direction-finding to locate the sources of TMR (ref. 7, and P. J. Christiansen, M. P. Gough and J. Etcheto, personal communication), and the sources were found to lie on the dayside plasmopause and on the morning magnetosheath⁷. It was also established that the radiation is significantly, if not predominantly, propagating in the left-hand ordinary (L-O) magnetoionic mode.

The hypothesis for the production of TMR considered here^{13,14} is that electrostatic waves close to the upper-hybrid-resonance (UHR) frequency $f_{UHR} = (f_{pe}^2 + f_{ce}^2)^{1/2}$ are the source of TMR. These intense UHR waves, which exist in a warm plasma, are tightly confined to within a few degrees of the magnetic equator at, and beyond, the Earth's plasmopause⁸. They are probably produced by a warm loss-cone component of the energetic electron distribution as outlined by Ronnmark *et al.*¹⁵. By virtue of their propagation into a slowly varying plasma density, it is suggested that their wave numbers decrease such that they access directly the cold-plasma Z-mode branch¹⁶ of the dispersion relation, which is a natural extension of the warm-plasma branch. Ray-tracing of Z-mode waves show how subsequent access can be gained by these waves to magnetoionic window points^{17,18}; these allow the direct coupling of energy to L-O mode waves which then propagate to lower density regions. These L-O waves are found to be beamed away from the geomagnetic equator by the presence of the local magnetic field.

Jovian kilometric radiation

The Voyager spacecraft observed radio emissions in the frequency range 10 kHz–1 MHz, emanating from Jupiter's inner magnetosphere¹⁻⁴. The emissions seem to suffer a distinct latitudinal beaming such that beyond the orbit of Io, a shadow zone exists with virtually no radiation seen within $\pm 10^\circ$ of the zenomagnetic equator¹⁹. Furthermore, the latitudinal extent of the shadow zone increases with increasing wave frequency, so that the higher frequencies are beamed to higher magnetic latitudes. Polarization measurements⁴ have shown that radiation seen north of the shadow zone, that is when the jovian magnetic north pole is tilted towards the spacecraft, is left-hand polarized, whereas that observed in the southern magnetic hemisphere is right-hand polarized; polarization is defined with respect to the direction of the wave-normal at the observer.

Two distinct and separate locations have been suggested as being possible source regions of JKR. The first source is assumed to lie along auroral magnetic field lines linked to the Io torus²⁰. The torus is a region of enhanced plasma density believed to extend around Jupiter, threaded by the orbit of Io, and therefore

* Present address: British Antarctic Survey, NERC, Madingley Road, Cambridge CB3 0ET, UK.

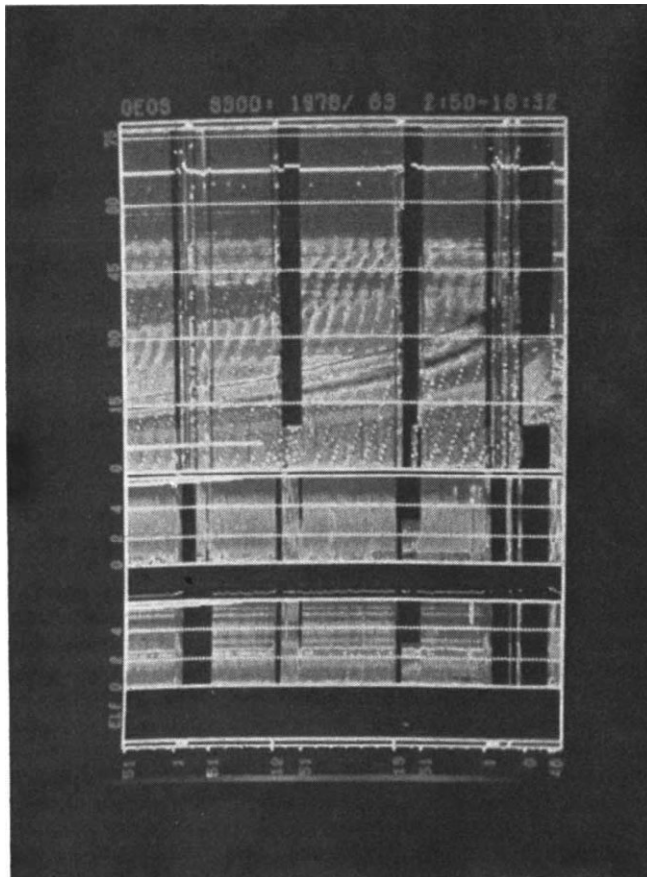


Fig. 1 Diagrammatical representation of wave spectrogram obtained on GEOS 1 on 4 March 1978. Terrestrial myriametric radiation appears as inclined striations due to the preferential direction of arrival of the waves. Intense upper hybrid waves occur at ~ 30 kHz after 04.00 UT.

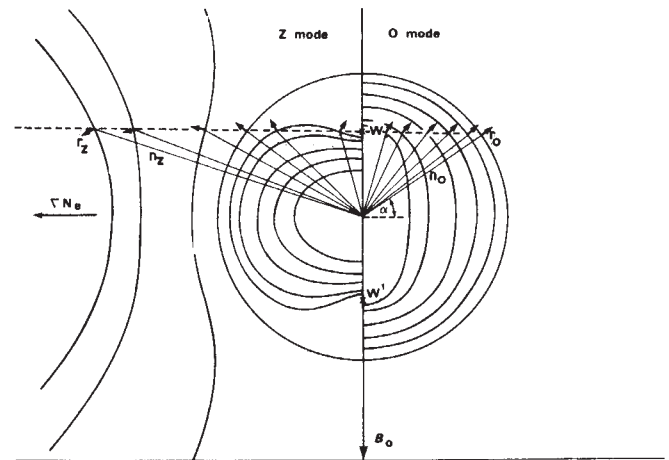
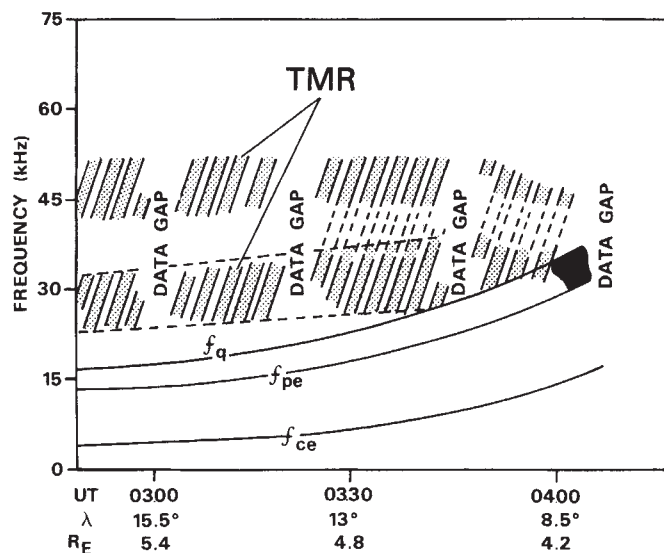


Fig. 2 Refractive index curves for the Z and O modes assuming that the gradient of electron density ∇N_e is perpendicular to the magnetic field B_0 as at the magnetic equator. A Z wave with initial wave normal n_z and corresponding ray r_z passes through the coupling point W to become an O wave finally having wave normal n_0 and ray r_0 , which make angle α with the magnetic equator.

has shown that the equatorial shadow effect observed can be reasonably accounted for by this model, although the detailed extent of the shadow zone must depend on the assumed plasma density along auroral field lines. No specific mechanism has been suggested for wave emission at $2f_{pe}$, and as far as we know no unambiguous identification of $2f_{pe}$ emissions has as yet been made within the terrestrial and jovian magnetospheres.

The other possible source is the Io torus itself². To produce a shadow zone beyond the torus requires some sort of beaming mechanism, possibly inherent to the source. Another difficulty is an observation by Voyager 1, whose periapsis was at $5 R_J$, which allowed it to traverse the torus centred at $\sim 5.8 R_J$. Peak densities within the torus were found to correspond to $f_{pe} \sim 600$ kHz, whereas JKR extending up to ~ 1 MHz has been detected. These results suggest that the Io torus may be unable to support emission at $f_{pe} \sim 1$ MHz, the plasma frequency having been a possible candidate for the source emission. However, Voyager 1 probably did not pass through the region of peak density and Voyager 2, whose periapsis was at $10 R_J$, did encounter higher densities outside the torus than did Voyager 1 (ref. 4). Therefore plasma frequencies in the torus could exceed 1 MHz.

It is proposed here that JKR is created at, and beyond, the Io torus in a manner similar to that suggested for TMR. Within $\sim 23 R_J$ of Jupiter, the Voyager spacecraft observed durable, intense UHR waves which, as was found at Earth, were very tightly confined to within $\pm 2^\circ$ or less of the zenomagnetic equator⁹. However, much less is known about the Io torus than about the Earth's plasmopause and JKR may yield a significant insight into the characteristics of the torus at the magnetic equator.

Theory

If a region of large density gradients^{13,14} is located at the magnetic equator as at the plasmopause, then mode-coupling of Z-mode to O-waves could occur, and its consequences on the radiation can be illustrated by the refractive index surfaces in Fig. 2. The magnetic field B_0 is shown with the electron-density gradient, ∇N_e , assumed perpendicular to it as expected at the equatorial plasmopause. Z-mode refractive index surfaces have rotational symmetry about the magnetic field, but only one half is shown. The O-mode refractive index surfaces are also shown. The surfaces are for various plasma densities assuming the magnetic field to be constant; density in this refractive index domain increases towards the centre. The points W and W' are the window points through which Z waves can escape as O waves^{12,18}. Snell's law requires that, during propagation, $n \cos \theta$ remains constant, n being the modulus of the refractive index

lies close to the centrifugal and zenomagnetic equatorial planes. The auroral emission has been assumed to occur at $2f_{pe}$ in the L-O mode²⁰, and different frequencies would come from different altitudes above Jupiter, the higher frequencies being emitted from regions of higher densities closer to the planet. In this first theory, the Io torus acts merely as a passive obstacle to wave propagation, and ray-tracing in a model magnetosphere

vector in the direction of the wave normal, and θ its angle with respect to the magnetic field. Therefore, a line parallel to the density gradient through point W represents the locus of the tips of the vectors from n_z to n_o for the wave which couples from Z to O mode. The wave frequency, which is greater than the electron gyrofrequency, is assumed constant, the wave encountering the window when $f = f_{pe}$. Clearly, as the Z-O waves propagate through the coupling point to lower densities, the ray direction swings from r_z to r_o and tends towards an angle α with the equator, where α is arcs in $(f_{ce}/f_{ce} + f)^{1/2}$ (see ref. 17) and f_{ce} is the gyrofrequency at the window point, where $f = f_{pe}$. A similar ray will exist for window point W^1 , and if mode coupling is the mechanism for TMR and JKR emission then the radiation would be beamed away from the magnetic equator at an angle α . Figure 3 shows how α depends on f_{pe}/f_{ce} , and for typical plasmaspheric densities, this ratio is in the region of 10 yielding an α of $\sim 17^\circ$, although if the plasmapause moves to smaller radial distances such that f_{ce} increases, then α may be much larger. Clearly if there is a large density gradient such that f_{pe} is increasing more rapidly than f_{ce} , then the higher frequencies from that region will be beamed to lower latitudes, and if the gradients are small so that f_{ce} increases more rapidly than f_{pe} , then the converse is true. If the plasma density gradient is radial, the beams will lie in the magnetic meridian plane; if it is not radial, the beams will lie in the plane containing the density gradient and magnetic field vectors.

TMR ray paths

The terrestrial magnetospheric model used for the ray tracing has been described elsewhere²¹, although the plasmasphere model has been updated using information from GEOS 1.

Figure 4 shows the ray path in the magnetic meridian plane at $4R_E$ of the Z-mode wave which couples to the O-mode, and which was in fact traced back from the coupling point. The arrows on the ray path indicate the directions of the wave normal. At S the wave normal is approximately parallel to the density gradient and perpendicular to the magnetic field. (The computations were stopped at S because the wave moves very slowly and the computation time becomes excessive.) The wave

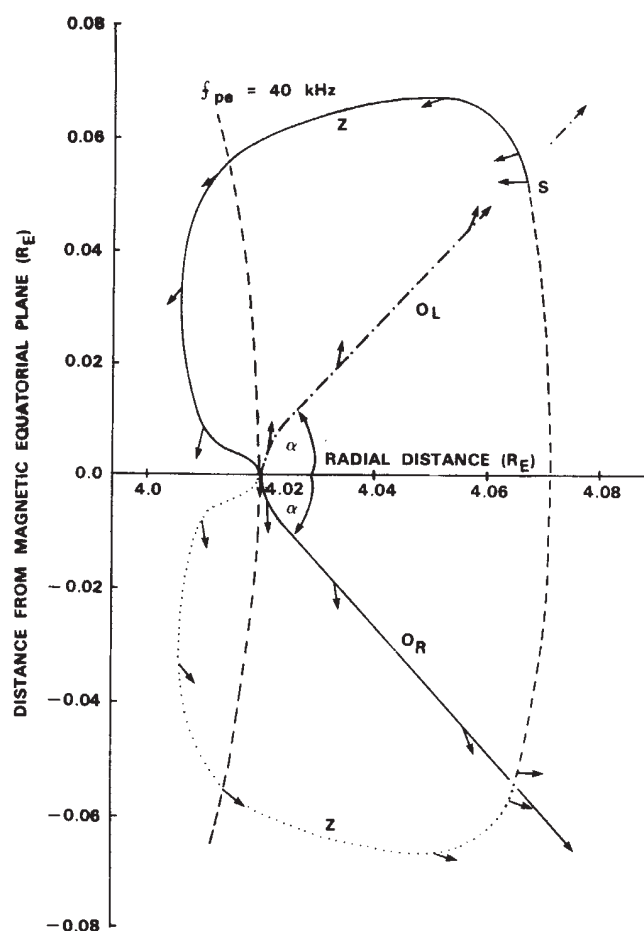


Fig. 4 The ray path in the magnetic meridian plane of a 40-kHz wave which starts at S with its wave normal, depicted by the small arrow, approximately perpendicular to the magnetic field and parallel to the density gradient. Mode-coupling occurs at the magnetic equator and the escaping O_R wave (solid line) propagates to free space at angle α to the magnetic equator; the contour $f_{pe} = 40$ kHz is shown. Reversing the wave-normal of the Z wave in the southern hemisphere allows coupling to occur at the other window to produce the O_L ray as shown in Fig. 2.

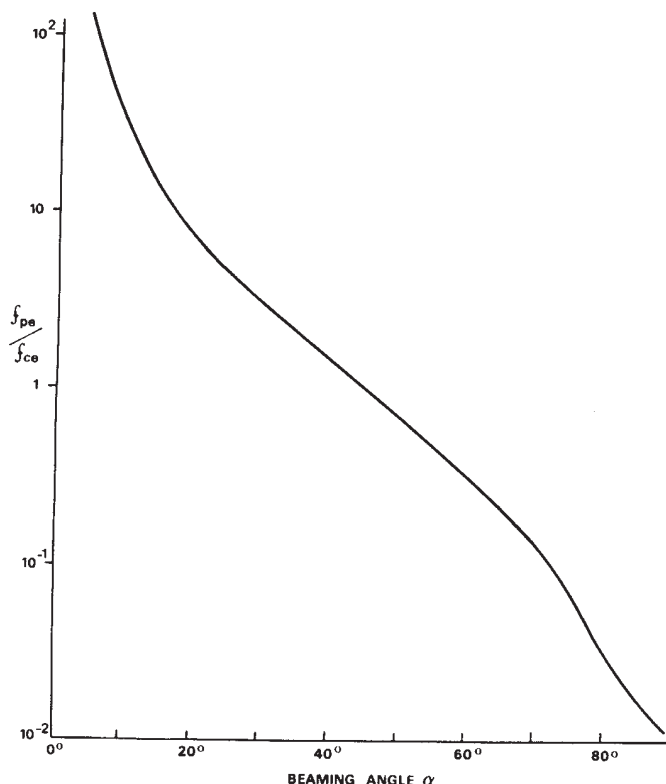


Fig. 3 The variation of the beaming angle α as a function of the ratio of electron plasma frequency to gyrofrequency, f_{pe}/f_{ce} .

normal in this direction is very important because electrostatic UHR waves can propagate¹⁵ without damping only with their wave normals close to perpendicular to the magnetic field. Therefore if the warm-plasma electrostatic waves propagate to link on to the cold-plasma Z-mode, it might be expected that at that point (S) the Z wave-normal will be as shown. The possible significance of such Z wave-normals, parallel to the density gradient, had been pointed out previously²². The Z ray eventually reaches the coupling point with its wave normal parallel to the magnetic field, and the O ray which couples from it moves to lower plasma densities at an angle α to the geomagnetic equator. This ray, O_R , will be observed to be right-hand polarized by a satellite; the ray O_L from the other window point will be observed to be left-hand polarized. Unfortunately, no polarization measurements of TMR have yet been performed.

It is significant that had the initial Z wave normal at S been such that the window had been just missed, the Z ray would propagate symmetrically about the equator, becoming electrostatic again, with its wave normal reversed, and possibly yielding its energy back to the charged particles on passing once more into the equatorial interaction region. Note that the latitudinal excursion of the Z ray which couples out is only $\sim \pm 1^\circ$ from the magnetic equator, which was found to be approximately the extent of the intense UHR waves seen on GEOS and other spacecraft.

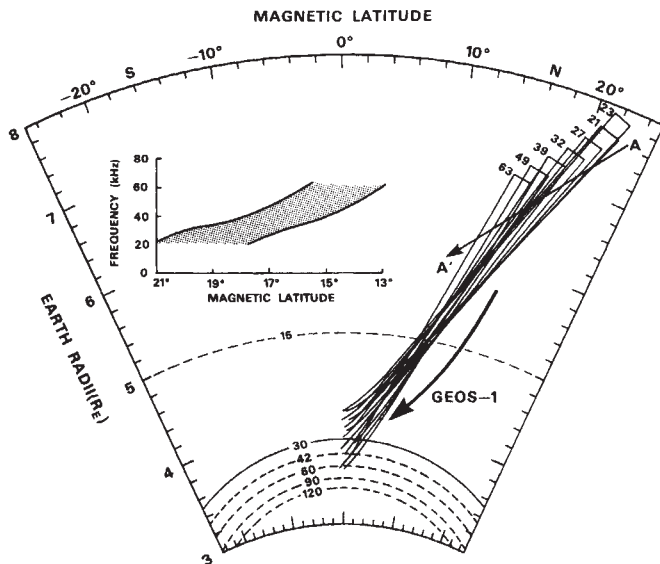


Fig. 5 Beams of O radiation for different frequencies assuming each has coupled out at the levels corresponding to $f = f_{pe}$. The plasma frequency contours are labelled 15–120 kHz. The GEOS 1 track is shown and the inset shows the type of spectrogram observed by a satellite moving along track AA'.

If one assumes that the window points are distributed $\pm 0.5^\circ$ about the Equator, then the resulting beams of TMR in the Northern Hemisphere are as shown in Fig. 5; identical beams exist in the Southern Hemisphere. The beams for a whole range of frequencies f are illustrated, each being assumed to have penetrated the window at the relevant level where $f = f_{pe}$. The track of GEOS 1 corresponding to the event on 4 March 1978 is also shown. The plasmopause in the model used in the ray-tracing needs to be moved inwards to slightly smaller radial distances. For a given frequency the latitudinal beam angle α will then increase because f_{ce} increases, and f_{pe}/f_{ce} decreases. That the TMR above 45 kHz in the GEOS-1 spectrogram remains virtually unchanged throughout the time of observation indicates that the satellite was moving within and parallel to the beam at those frequencies. A satellite moving along track AA' in Fig. 5 would record a spectrogram similar to that shown in the inset. However, much more complicated spectrograms can be envisaged depending on the satellite track chosen. (An example of a TMR spectrogram in which the frequency decreases as the satellite moves inwards towards the plasmopause may be seen in Fig. 7 of ref. 23.) The frequency bandwidth of the TMR received at a given satellite position is indicative of the width of the beam, and the assumed $\pm 0.5^\circ$ latitudinal extent of the source seems to be approximately that required to explain the GEOS 1 spectrogram.

JKR ray paths

Ray-tracing has also been performed in a model jovian magnetosphere assuming that UHR waves at the Io torus and beyond are responsible for JKR. The model magnetosphere, detailed elsewhere²⁴, is largely similar to that of the Earth's taking into account the opposite polarity of the magnetic dipole, the different magnetic field strength and the existence of the Io plasma torus. If the Io torus has a low-curvature rounded outer edge at the equator, the O waves, which come from the proposed coupling process, will be beamed in a similar manner to those from the Earth's plasmopause. Because Jupiter's magnetic dipole is inverted compared with that of the Earth's, the beam which would be observed in the northern magnetic hemisphere would be left-hand polarized as seen by a spacecraft. This is opposite to what was observed by the Voyager spacecraft, and to

adjust the model to account for this, a distended torus is necessary, as shown in Fig. 6. The contours of constant plasma density are labelled with their corresponding plasma frequencies and the ray paths are shown for different wave frequencies, assuming that mode-coupling has occurred at $\pm 0.5^\circ$ latitude. Provided the torus distension tilts with the tilting magnetic field, a spacecraft located at a given centrifugal latitude will observe tapered emissions of the type seen by the Voyager spacecraft; as the magnetic latitude of the spacecraft increases, higher frequencies are observed, and a shadow zone exists within $\sim \pm 10^\circ$ of the magnetic equator.

This beaming of JKR depends on the shape of the torus distension at the equator and hence a detailed analysis of JKR could yield valuable information on those torus regions not visited by the Voyager spacecraft. The rocking of the torus by the magnetic field is to be expected as the Io torus is believed to contain anisotropic charged particle distributions favouring large pitch angles, and any gradual tilting of the magnetic field would cause the particles to be moved bodily with it.

Also, if there were propagation through the other window point, then in Fig. 6 the result would appear as very narrow beams (dashed lines) going to high latitude as shown for a source at $8 R_J$. Whether these exist probably depends on the ray-paths of the corresponding Z-mode waves. Figure 7 shows the ray paths of Z waves which would couple through the two windows. The coupling is assumed to occur at $8 R_J$ at latitude 0.5° ; also shown are the resulting O-mode rays, O_L going to a low latitude and O_R to a high latitude. Data from the Voyager Planetary Radio Astronomy experiment should confirm whether the high-latitude beams exist. As Voyager 2 moved inwards towards the torus the plasma wave experiment observed that at about $46.5 R_J$ the Northern source split into two³, and this effect may be produced by a single narrow beam which is cut twice by the spacecraft—once as its magnetic latitude increases and once as it decreases. On Voyager 2 the strong JKR at ~ 17 –56 kHz disappeared at $\sim 33 R_J$ indicating that the source region for these frequencies had been passed, and clearly there was a fairly large density gradient at $\sim 33 R_J$ such that the plasma frequency

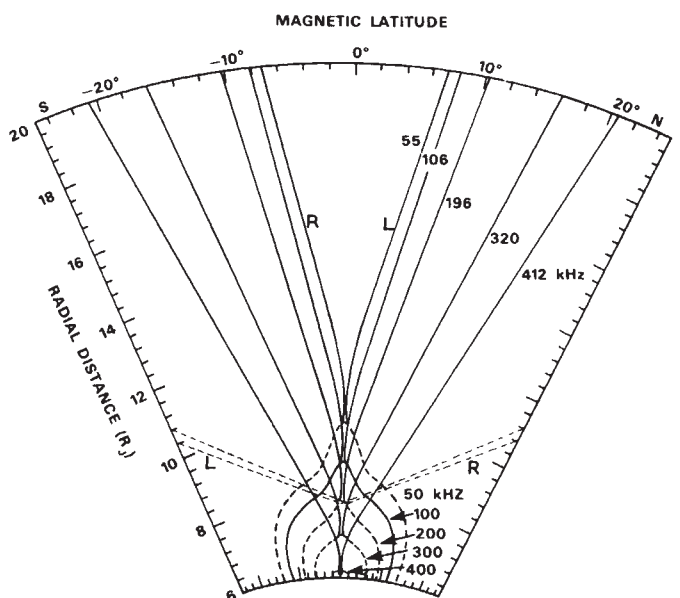


Fig. 6 The low-latitude limits of the 'main' beams of jovian kilometric radiation showing the latitudinal dependence on frequency. The plasma density contours representing the distended Io torus are labelled according to plasma frequency (50–400 kHz). The narrow beams to high latitudes are from propagation through the second coupling point (see text), the L and R depicting left-hand and right-hand polarization respectively as observed by a spacecraft.

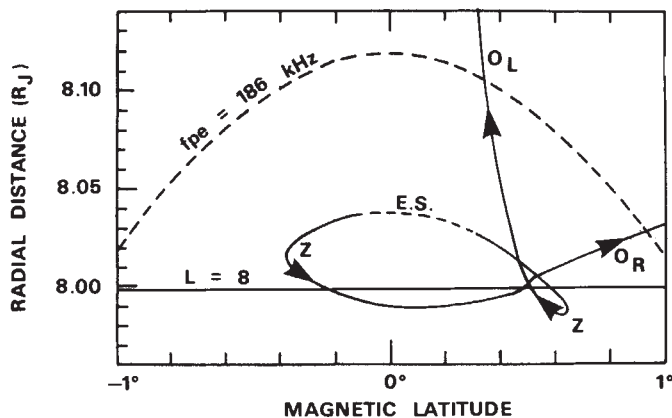


Fig. 7 Details of ray paths close to the coupling region. ES shows where the upper hybrid waves are electrostatic. One Z ray propagates to -0.4° magnetic latitude before it reaches the coupling point at $+0.5^\circ$ latitude, subsequently propagating as the O_R ray going to high latitude. The other Z ray turns at $+0.65^\circ$ latitude and becomes the O_L ray. For reference the $L = 8$ magnetic field line is shown as is the plasma frequency contour corresponding to 186 kHz. The wave frequencies are all 196 kHz.

increased rapidly from 17 kHz to above 56 kHz in a relatively short distance. Perhaps narrow-band JKR sometimes observed as a stronger emission within wide-band JKR has its source in such large density gradients, which increase the window width, thereby allowing the escape of more Z-mode energy¹⁸. There has been some controversy as to the identification of the UHR frequency on Voyager 2 (refs 3, 4) and, on the basis of the theory proposed here, the 56 kHz UHR frequency was at $\sim 30 R_J$ and the UHR frequency at $\sim 10 R_J$ was probably ~ 193 kHz. However, in a highly distended torus latitudinal variations are an important factor when interpreting the *in situ* density estimates made by the Voyager experiments.

Discussion

The theory that intense electrostatic UHR waves coupling to L–O waves through the Z-mode are the source of JKR and TMR can account for many characteristics of the radiations. Detailed ray tracing of the electrostatic UHR waves in a realistic magnetospheric model and calculations of the width of the coupling windows are underway to confirm some features of the hypothesis, and to calculate the efficiency of the processes. The powers of TMR and JKR previously estimated on the basis of isotropic emission will have to be correspondingly reduced. Limiting the symmetrical Z-mode ray paths to $\pm 1^\circ$ of the magnetic equator is believed to be significant and may be a crucial factor in the instability and in the energy exchange between waves and particles at the equator. That the UHR waves can access the Z-mode branch may indicate where exactly on the warm dispersion branch the UHR waves are produced²⁵.

Note also that the confinement of UHR waves and associated particle distributions^{8,26} may allow the first determination of the actual position of the magnetic equator, which would be very important within the context of other types of equatorially produced emissions, such as whistler-mode 'chorus'²⁷.

A more detailed analysis of Voyager wave data should yield much new information on the Io torus. For example, the theory proposed here would predict that lower-frequency long-lived features of JKR should exhibit less co-rotation than higher frequencies because the former have their source at larger radial distances.

An important consequence of the theory, because the propagation windows are only a few degrees wide in practice¹⁸, is that JKR and TMR will be beamed away from the source region such that the rays lie very close to the plane containing the magnetic field and the density gradient vectors. This provides a powerful diagnostic technique allowing one to map the shapes of the plasmopause and Io torus in their respective equatorial planes by performing direction-finding measurements on the radiation. JKR shows long-lasting co-rotating features which reappear at approximately the same longitude for many jovian rotations. The Earth's plasmopause may co-rotate without large changes in its shape provided that one is not in regions affected temporally by large substorm electric fields^{28,29}. High-resolution direction finding (DF) measurements are being performed using GEOS and ISEE data and the results will be reported later.

These comments on the plasmopause source are equally valid for the other source of TMR—the morning magnetosheath. Gurnett *et al.*³⁰ have reported ISEE observations of narrow band emissions near the electron plasma/upper hybrid frequencies in this region of the magnetosheath and I suggest that the energy in these waves also escapes as L–O mode radiation in a manner similar to that proposed for the plasmopause source. The magnetosheath magnetic field and plasma flow velocities are continuously changing depending on the direction of the interplanetary magnetic field and solar wind velocity and it is more difficult to determine the shape of the magnetosheath on the basis of the DF measurements. However, in stable conditions and with high time-resolution DF data it may be possible to map the magnetosheath and to see the propagation of irregularities and large-scale waves within it.

Finally, I re-emphasize that the determination by spacecraft of the local plasma density from the sharp low-frequency cut-off of TMR should be treated with caution. Clearly the cut-off is often due to beaming effects and bears no relation to the local plasma frequency at the satellite except in that it provides an upper limit to it. This may be equally valid for the low frequency jovian emissions tentatively identified as non-thermal continuum, which were detected by Voyagers 1 and 2 on their outbound passes^{1,3}, and which were observed also to have a shadow zone at the magnetic equator.

I thank K. G. Budden, P. Christiansen, J. Etcheto, C. Goertz, P. Gough, R. Grard, D. Gurnett, R. Helliwell, R. Herring and M. J. Rycroft for stimulating discussions, R. M. Laws for support, and E. Gershuny, M. Trischberger and C. Boucher for computing assistance.

Received 24 July; accepted 12 September 1980.

1. Scarf, F. L., Gurnett, D. A. & Kurth, W. S. *Science* **204**, 991–995 (1979).
2. Warwick, J. W. *et al. Science* **204**, 995–998 (1979).
3. Gurnett, D. A., Kurth, W. S. & Scarf, F. L. *Science* **206**, 987–991 (1979).
4. Warwick, J. W. *et al. Science* **206**, 991–995 (1979).
5. Gurnett, D. A. *J. geophys. Res.* **80**, 2751–2763 (1975).
6. Brown, L. W. *Astrophys. J.* **180**, 359–370 (1973).
7. Shaw, R. R. & Gurnett, D. A. *J. geophys. Res.* **78**, 8136–8149 (1973).
8. Gough, M. P., Christiansen, P. J., Martelli, G. & Gershuny, E. J. *Nature* **279**, 515–517 (1979).
9. Kurth, W. S., Barbosa, D. D., Gurnett, D. A. & Scarf, F. L. *Geophys. Res. Lett.* **7**, 57–70 (1980).
10. Jones, D. *IEEE Trans. Geosci. Electron.* **GE-12**, 9–18 (1974).
11. Jones, D. *Annls. Télécommun.* **35**, 187–195 (1979).
12. Jones, D. *Space Sci. Rev.* **22**, 327–332 (1978).
13. Jones, D. *Nature* **260**, 686–689 (1976).
14. Jones, D. *The Scientific Satellite Programme During the International Magnetospheric Study*, 281–291 (Reidel, Dordrecht, 1976).
15. Ronnmark, K., Borg, H., Christiansen, P., Gough, P. & Jones, D. *Space Sci. Rev.* **22**, 401–417 (1978).
16. Oya, H. *Planet. Space Sci.* **22**, 687–708 (1974).
17. Budden, K. G. *Radio Waves in the Ionosphere* (Cambridge University Press, 1966).
18. Budden, K. G. *J. atmos. terr. Phys.* **42**, 287–298 (1980).
19. Gurnett, D. A., Kurth, W. S. & Scarf, F. L. *Science* **206**, 987–990 (1979).
20. Green, J. L. & Gurnett, D. A. *Geophys. Res. Lett.* **7**, 65–68 (1980).
21. Jones, D. & Grard, R. J. L. *The Scientific Satellite Programme During The International Magnetospheric Study* 293–302 (Reidel, Dordrecht, 1976).
22. Jones, D. *J. Geomagn. Geoelectr.* **30**, 291–293 (1978).
23. Gurnett, D. A., Anderson, R. R., Scarf, F. L., Fredericks, R. W. & Smith, E. J. *Space Sci. Rev.* **23**, 103–122 (1979).
24. Jones, D. *Proc. COSPAR Symp. on Planetary Magnetospheres*, Helsinki (in the press).
25. Barbosa, D. D. *J. Geophys. Res.* **85**, 2341–2345 (1980).
26. Wrenn, G. L., Johnson, J. F. & Sojka, J. J. *Nature* **279**, 512–514 (1979).
27. Helliwell, R. A. & Crystal, T. L. *J. geophys. Res.* **78**, 7357–7371 (1973).
28. Carpenter, D. L. *J. geophys. Res.* **71**, 693–709 (1966).
29. Carpenter, D. L., Park, C. G. & Miller, T. R. *J. geophys. Res.* **84**, 6559–6563 (1979).
30. Gurnett, D. A. *et al. J. geophys. Res.* **84**, 7043–7058 (1979).

Greenland ice sheet evidence of post-glacial volcanism and its climatic impact

C. U. Hammer, H. B. Clausen & W. Dansgaard

Geophysical Isotope Laboratory, University of Copenhagen, Haraldsgade 6, DK 2200 Copenhagen N, Denmark

Acidity profiles along well dated Greenland ice cores reveal large volcanic eruptions in the Northern Hemisphere during the past 10,000 yr. Comparison with a temperature index shows that clustered eruptions have a considerable cooling effect on climate, which further complicates climatic predictions.

HISTORICALLY, man has often been directly influenced by volcanic eruptions: for example Mont Pél  (Caribbean, 1902), Krakatoa (Indonesia, 1883), Tambora (Indonesia, 1815), Laki (Iceland, 1783), Vesuvius (Italy, AD 79), and Thera (Santorini, ?). Other great eruptions have been observed only by the associated optical phenomena in the atmosphere caused by volcanic aerosols, for example AD 1601–02 (ref. 1) and 44 BC (ref. 2). But, before AD 1600 most volcanism was not recorded, and not even noticed outside the regional areas that suffered from the direct impact of the eruptions.

However, indirect impacts may be expected, because great volcanic eruptions inject huge amounts of silicate micro-particles and acid gases into the stratosphere. The main effect of the silicates is probably to heat up the stratosphere by absorption of solar radiation^{3,4}, but they disappear by settling during the first few months^{5,6}. At the same time a sulphate aerosol, mainly sulphuric acid, is building up from SO₂ to become the dominant aerosol^{6,7}, the main effect of which seems to be cooling the lower troposphere by back scattering of solar radiation^{3,4}. Most of this information has been obtained by directly monitor-

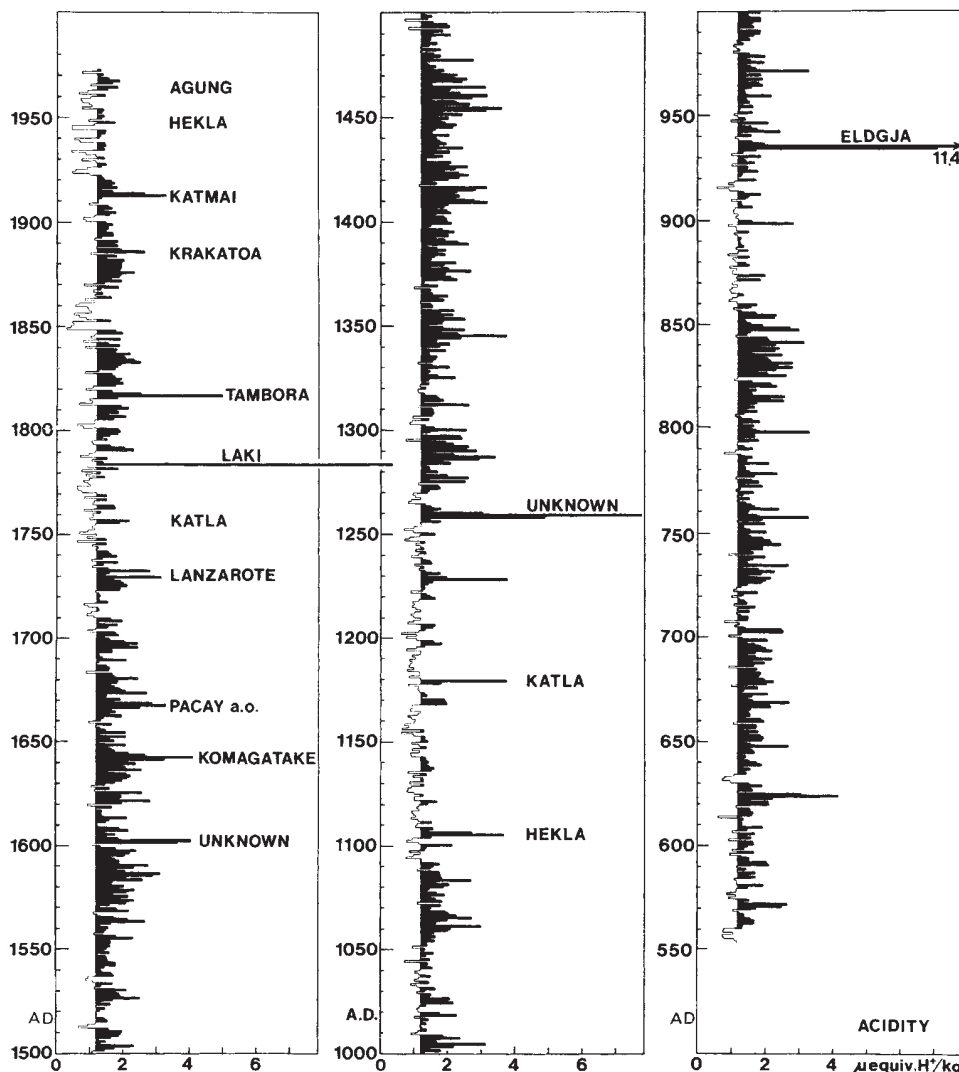


Fig. 1 Mean acidity of annual layers from AD 1972 to 553 in the ice core from Cr te, Central Greenland. Acidities above the background, 1.2 ± 0.1 $\mu\text{equiv. H}^+$ per kg ice, are due to fallout of volcanic acids, mainly H₂SO₄, from eruptions north of 20 S. The ice core is dated with an uncertainty of ± 1 yr in the past 900 yr, increasing to ± 3 yr at AD 553 (ref. 10), which makes possible the identification of several large eruptions known from historical sources, and the accurate dating of the Icelandic Eldgja eruption shortly after the settlement that was completed AD 930.

Table 1 Magnitudes (in terms of global acid fallout) of some volcanic eruptions north of 20° S, roughly estimated from the acidity of annual layers in ice cores from Greenland

Volcano	Location	Year of eruption	Volcanic acid fallout in Greenland (H ₂ SO ₄ +HX) kg km ⁻²	Latitude correction factor* (1/F)	Global acid fallout (H ₂ SO ₄ +HX) (× 10 ⁶ tons)
Crête					
Agung	Indonesia	1963 AD	9	3	20
Hekla	Iceland	1947	6	1	5
Katmai	Alaska	1912	37	1	30
Krakatoa	Indonesia	1883	21	3	55
Tambora	Indonesia	1815	58	3	150
Laki	Iceland	1783	116	1	100
Lanzarote	Canary Isl.	1730–36	36	2	60
Krafla (Myvatn fires)	Iceland	1724–30	63	1	55
Unknown	?	1666?	?	?	?
Pacaya	Guatemala	1664	60	3(2)	(100)
Awu	Indonesia	1641	?	3	?
Adiksa	Indonesia	1641	81	3(3)	(190)
Komagatake	Japan	1640	?	2	?
Unknown	?	1600/01	61	1?	(50)
Raudubjallar	Iceland	1554	23	1	20
Unknown	?	1257/58	128	1?	(110)
Hekla 1	Iceland	1104	51	1	45
Eldgjá	Iceland	934	189	1	165
Unknown	?	622/23	52	1?	(45)
Camp Century					
Unknown	?	50±30 BC	192	1?	(120)
Unknown	?	210±30	72	1?	(45)
Unknown	?	260±30	54	1?	(35)
Hekla 3	Iceland	1120±50	99	1	60
Thera	Greece	1390±50	98	2	125
Hekla 4	Iceland	2690±80	96	1	60
Unknown	?	3150±90	255	1?	(160)
Mt Mazama	Oregon, USA	4400±110	156	2	200
Hekla 5/Thjórsá flow	Iceland	5470±130	90	1	60
Unknown	?	6060±140	119	1?	(75)
—	?	6230±140	102	1?	(65)
—	?	7090±160	79	1?	(50)
—	?	7240±160	124	1?	(80)
—	?	7500±160	51	1?	(35)
—	?	7640±170	(412)	1?	(260)
—	?	7710±170	69	1?	(45)
—	?	7810±170	73	1?	(45)
—	?	7910±170	95	1?	(60)

* See ref. 8.

ing the atmosphere in the years after the few great eruptions that have occurred in the past 20 yr, but this period of observation is obviously too short to allow any conclusion as to a possible persistent climatic effect of volcanic activity.

In the Northern Hemisphere (north of 20° S), the Greenland ice sheet is a rich source of information on past volcanism. This is not due to deposition of volcanic micro-particles⁸, as these are apparently too large and/or too few to reach the high-altitude regions of the Greenland ice sheet in amounts that significantly raise the (continental) dust concentration background. However, volcanic acids in snow layers deposited shortly after a large volcanic eruption can be detected—as elevated specific conductivities measured on melted ice samples⁹, or as elevated acidities revealed by an electric current through the solid ice, when pulling a simple pair of electrodes (voltage difference ~1,250 V) along a clean cut surface of an ice core⁹. We used the latter method because it is fast, non-destructive and more selective, the current being almost entirely determined by the protons from the strong acids.

The Crête acidity record

A 404-m long ice core from Crête, Central Greenland (71° N, 37° W), was recovered in 1974 by US Cold Regions Research Engineering Laboratory under the American–

Danish–Swiss Greenland Ice Sheet Program (GISP). It has been dated to an accuracy of ±1 yr through the past 900 yr and to ±3 yr at the oldest annual layer deposited AD 553 (ref. 10).

Electrical measurements on the solid ice provide the acidity record shown in Fig. 1, except for the upper porous part of the core (back to AD 1840), where the acidity values are based mainly on pH measurements on melted samples. Some of the very low values recorded since AD 1840 may be due to the difficulty of making a proper correction for CO₂-induced H⁺ in water of low acidity. The same reason may have caused a few small signals to be omitted in the uppermost part of the record.

Low-latitude eruptions generally show up with a time lag of ~1 yr due to the long travel time of the aerosols before deposition in Greenland. Some of the easily identified eruptions are denoted in Fig. 1. The layers of highest acidity coincide with internal reflection horizons in radio-echo sounding^{8,11} and are thus recognizable all over Greenland.

The increase of the acid concentration in ice layers containing debris from a high northern latitude eruption can be used to estimate the global H⁺ fallout by applying the same latitudinal distribution pattern observed for ⁹⁰Sr fallout from high northern latitude bomb tests. In 1962, the global fallout of ⁹⁰Sr (mainly from the Soviet tests in 1961) was 1.2 MCi which was 8.6 × 10⁸ and 6.4 × 10⁸ times the fallout per km² at Crête and Camp Century, respectively. Applying this factor to the time–

integrated fallout of volcanic acids per km^2 from a given eruption gives the magnitude of the eruption in terms of acid-gas injection into the stratosphere. The fallout from mid-latitude (20° – 50° N) or low-latitude (20° S– 20° N) eruptions should be multiplied by factors of 2 or 3, respectively, to account for the spread of the eruption cloud during the travel to high northern latitudes, judged from fallout data (more extensive than in ref. 8) on debris from bomb tests outside the polar region. Magnitude estimates on several eruptions listed in Table 1 could easily be wrong by a factor of 2 for several reasons. For example, they may be underestimated, because some acid may have been neutralized before deposition, and in the case of Tambora, a higher fraction of the acid gases than during the explosive Krakatoa eruption may have remained in the troposphere where they were washed out at low latitudes. On the other hand, a high tropospheric wind pattern, favourable for deposition of volcanic acids in Greenland, may lead to an overestimation, particularly in the Icelandic eruptions.

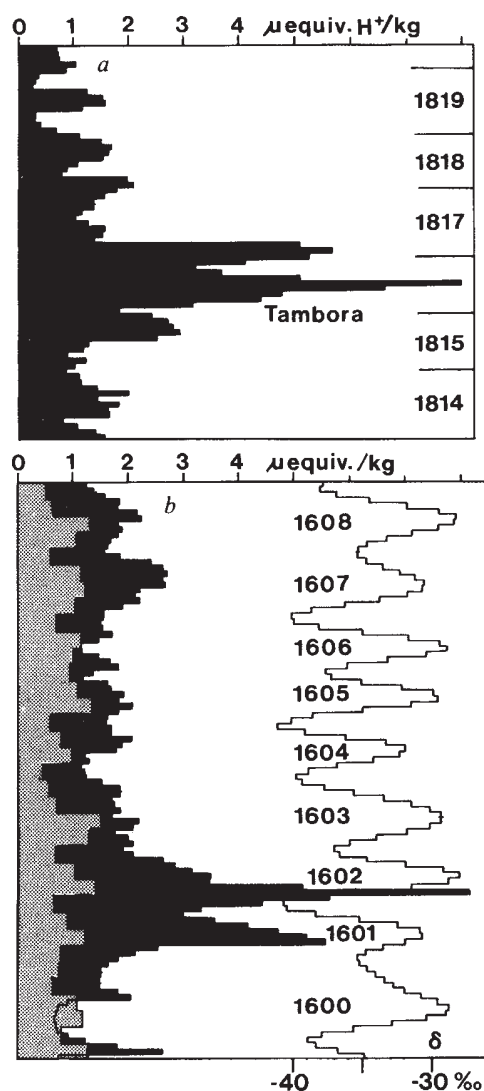


Fig. 2 The acid fallout (black curves; scale on top) from two long volcanic eruptions, found in annual layers at Crête. *a*, The Tambora eruption in Indonesia began AD 1815, and the fallout in Greenland peaked in 1816, "the year without a summer" (ref. 12). *b*, The $\delta^{18}\text{O}$ curve to the right shows that the fallout from an eruption in AD 1600 or early 1601 at an unknown locality lasted to the end of 1602. The NO_3^- concentration (shaded curve to the left) varies seasonally, but it was not above normal shortly after the supernova 1604 (see ref. 16).

Most of the eruptions recorded in the Crête core (Fig. 1) and listed in the upper part of Table 1 will now be considered individually.

Krakatoa (1883, 6° S) and Tambora (1815, 8° S) are revealed by relatively low signals due to the long distance to Greenland. Figure 2*a* shows that the acid fallout from the Tambora eruption peaked in mid 1816, "the year without a summer"¹².

Laki (1783) and Eldgjá (tenth century) are the two highest signals in the record; revealing the two greatest eruptions in Iceland since the landnam (settlement) of 875–930, perhaps since the last glaciation. The fallout from Laki (1783) has been detailed elsewhere⁸. The exact year of the Eldgjá eruption does not appear in Icelandic annals, but according to Thorarinsson (personal communication, and ref. 13) the eruption took place some time during or shortly after the landnam, although not later than AD 950. The eruption is identified (at AD 934 ± 2) in the Crête record with high certainty, because (1) the eruption was of the same magnitude as Laki (1783), at least 10 km^3 of lava, which is more than any other lava extrusion on Earth in historical time, and (2) there is no other signal comparable to that of Laki (1893) in the said period.

Hekla (1104) is the greatest Hekla eruption (H1) in Icelandic historical time, that is, in the past thousand years. The annals generally date it at 1104, and "the winter of sand-rain" 1104–05 is discussed which suggests that the eruption began late in 1104. In a detailed acidity record⁹ the acidity increases upward from the top of the annual layer previously dated at 1104 by the δ method, and Fig. 1 shows an acidity peak in 1105, which is strong evidence that the time scale is correct at around AD 1100.

Most signals, particularly before AD 1600, are due to unknown eruptions somewhere in the Northern Hemisphere. The three strongest such signals are those in AD 1601–02, 1258–59 and 623–24 (see Fig. 1). The eruption corresponding to the first signal caused "a constant haze over southern Scandinavia in 1601. . . The Sun and Moon appeared 'reddish, faint and lacked brilliance' in central Europe all through the year 1601 and up to the end of July 1602" (ref. 1). Figure 2*b* spans the period 1599–1608, according to the dating based on counting seasonal δ -variations downward from top of the core. The δ -curve is first order corrected for smoothing by diffusion¹⁴. The acidity curve in the middle shows that the acid fallout lasted from the beginning of 1601 to the end of 1602. The location of the 1601–02 eruption cannot be deduced from these measurements, but the high peak of the signal suggests mid- or high latitudes (Alaska, Kamchatka?) because otherwise a low-latitude eruption causing this amount of fallout in Greenland would have to be of the Tambora class and could hardly have escaped the attention of the early colonialists. The NO_3^- concentration curve in Fig. 3 measured on melted samples¹⁵ shows that HNO_3 did not contribute to the elevated acidity signals in 1601 and 1602. Nor did the NO_3^- concentration increase above normal in 1604 and 1605 (the NO_2^- concentration was close to zero), which heavily influences the hypothesis of elevated nitrate production in the atmosphere during the supernova observed in 1604 (ref. 16), as will be discussed elsewhere¹⁷. Note that the NO_3^- concentration exhibits persistent seasonal variations (compare with the δ curve, which usually peaks in late summer), suitable for stratigraphic dating independent of volcanic eruptions. The NO_3^- summer peak may be due to the nitrate production in the atmosphere being determined by the intensity of the solar radiation, but the explanation is probably more complicated.

The third highest signal in the record spans three years and peaked in 1259. It cannot be ascribed to any Icelandic eruption, because the closest one in time (1262, probably a minor eruption of Katla; Thorarinsson, personal communication) is not within our uncertainty limits of about one year. The eruption must also have been large, because the integrated fallout is at least of the same order as that of Laki (1783).

Before AD 900 the volcanism revealed in the Crête record cannot be ascribed to any known eruptions. The highest signal on an annual basis is that of 623, and the absolute maximum acidity ($\sim 10 \text{ } \mu\text{equiv. H}^+ \text{ kg}^{-1}$) on a more detailed record occur-

red in 757. We doubt whether any of these signals were caused by the great eruption in the Yukon basin of Alaska some 1,400 yr ago, because an Alaskan eruption, estimated at 50% of Tambora (1815) in terms of ash production¹⁸, would be expected to cause considerably higher acid fallout in Greenland than indicated in Fig. 1. It is more likely that the Yukon eruption will show up if the acidity record is extended beyond AD 553—the year of deposition of the oldest layer in the 404-m long Crête core.

The Camp Century acidity record

The Camp Century, north-west Greenland (77° N, 61° W), ice core is the only one reaching from the surface to the bedrock in Greenland. The core is 1,390 m long, corresponding to 1,367 m of ice equivalent, and the deepest part of it is >100,000 yr old¹⁹. Unfortunately, there is a gap of ~510-yr between the oldest layer in the Crête ice core and the youngest part of the Camp Century core that is in sufficiently good physical condition for continuous acidity analyses (only a few outstanding events since AD 43 have been safely identified, for example, 1259 and Laki 1783, the fallout from which has approximately the same acidity as at Crête).

The dating of the Camp Century core cannot compare with that of the Crête core, although it has been dated by the δ method to better than $\pm 2\%$ through the interval¹⁰ 2,000–10,000 yr BP. Further improvement of the time scale over the past 4,000 yr has been obtained by using the AD 1259 and 1783 reference horizons as fixed points and also using all presently available data on the annual layer thickness λ , which decreases from 19 cm in 2,000-yr-old ice to 2 cm in 10,000-yr-old ice due to plastic thinning of the annual layers as they sink towards greater depths¹⁰.

A continuous acidity profile (one acidity mean value for each 10-cm increment) has been measured along the Camp Century

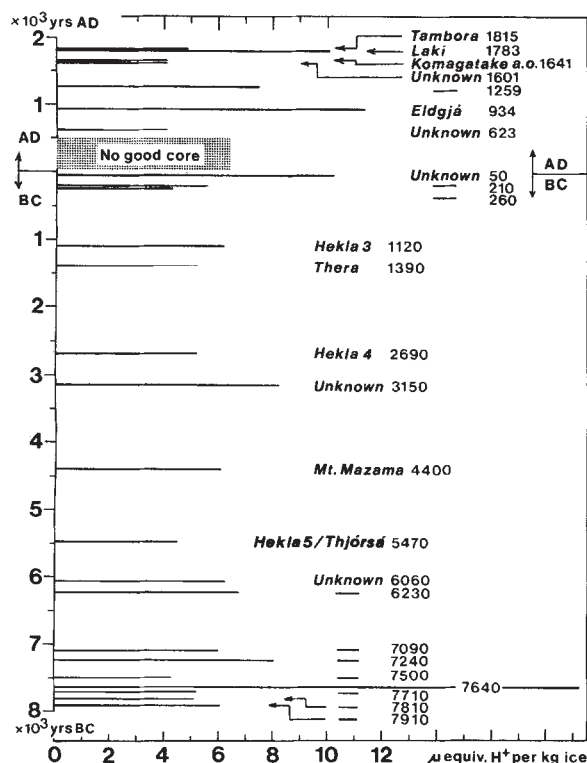


Fig. 3 Acidity signals exceeding $4 \mu\text{equiv. H}^+$ per kg ice in the past 10,000 yr, but for AD 44–552. Large eruptions have occurred most frequently in the earliest and in the latest millenium of the period (compare with ref. 30). Up to AD 43 the analyses were performed on the Camp Century ice core; since AD 553 on the Crête ice core.

ice core from a depth of 558 m of ice, corresponding to AD 43, to the bedrock. Very high acidity spikes were analysed back to 8,000 BC with a resolution higher than λ at the particular depths. For clarity Fig. 3 shows only the mean acidity of the most acid annual layer within each one of those signals that can be ascribed to a single violent eruption. This procedure probably reveals all the large, violent eruptions, whereas all the eruptions up to medium magnitude, as well as large long-lasting but less violent eruptions are deleted. For comparison, the Crête record in Fig. 1 has been treated in the same way and added to the top of Fig. 3. The approximate magnitudes of some of the eruptions in terms of acid-gas production are listed in Table 1, estimated by the same procedure as used on the Crête core.

The frequency of violent eruptions is clearly relatively low from 1500 to 7000 BC, in particular from 3500 to 5500 BC. Some of the most outstanding events are now discussed

50 $\pm$ 30 BC: The acid fallout lasted ~3 yr (compare with Fig. 4a) and must be due to one of the largest eruptions in the Northern Hemisphere since the last glaciation. There is no evidence of any great eruption in Iceland at this time. If Etna really had an eruption in the year of Caesar's death²⁰, 44 BC, it could explain Virgil's statement²¹ that "when Caesar died, the Sun felt pity for Rome, as it covered its beaming face by darkness, and the impious generation feared an eternal night". But, if the heavy fallout in Greenland were to be ascribed to Etna, the magnitude of the eruption and its impact on the Roman society would have been so great that there would undoubtedly be geological and historical evidence. A century later the Elder Plinius mentioned² the "feebleness of the Sun through nearly a year, when the dictator Caesar was killed", which suggests there was a widespread aerosol from an unusually large eruption somewhere else.

210 $\pm$ 30 or 260 $\pm$ 30 BC: Might reveal the Younger Laxar-lava (Iceland; Thorarinsson, personal communication).

1100 $\pm$ 50 BC: The greatest post-glacial eruption of Hekla (H3) has been radiocarbon dated at 950 ± 130 BC (ref. 22), that is not significantly different from 1100 ± 50 BC. The estimated total volume of freshly fallen tephra was 12 km^3 (ref. 23), corresponding to some 30% of the material extruded by Laki (1783).

1390 $\pm$ 50 BC: (Compare details in Fig. 4b). This is the only signal exceeding $2.6 \mu\text{equiv. H}^+$ kg^{-1} between 1100 and 2700 BC, and we therefore interpret it as being due to the large eruption of Thera (Santorini) in the Aegean Sea, which is generally agreed to have been of the same magnitude as that of Tambora (1815). The tephra production has recently been estimated at more than 28 km^3 (13 km^3 of dense rock equivalent)²⁴. This unusually large eruption has been radiocarbon dated at 1720 ± 50 BC on the calibrated radiocarbon scale (ref. 25, and H. Tauber, personal communication). However, archaeological evidence from the excavation of the Minoan settlement near Akrotiri from Santorini strongly suggests²⁶ that the island was inhabited at least up to 1500 BC judging by Egyptian pottery style chronology; it was apparently abandoned shortly before the eruption, and in good order because no valuables have been found nor people killed by the heavy ash fall (10–40 m). The discrepancy between the datings may be partly explained if the organic material used for the radiocarbon dating were partly built up by radioactively dead CO_2 exhausted from the volcano before the eruption (an effect which has been observed recently²⁷). Our dating around 1400 BC supports Marinatos' theory²⁸ of a causal connection between the Thera eruption and the decline of the Minoan civilization centred on the island of Crete. The dating can be further improved to ± 10 yr, if and when a deep Central Greenland ice core becomes available.

2700 $\pm$ 70 BC: Agrees with the radiocarbon dating 2550 ± 120 BC of the Hekla eruption H4, which is estimated to have been almost as great as H3 (ref. 23).

3250 $\pm$ 80 BC: This signal is difficult to reconcile with any known giant eruption, but the one at

4400 $\pm$ 100 BC: Agrees with Mt Mazama, Oregon, the radio-

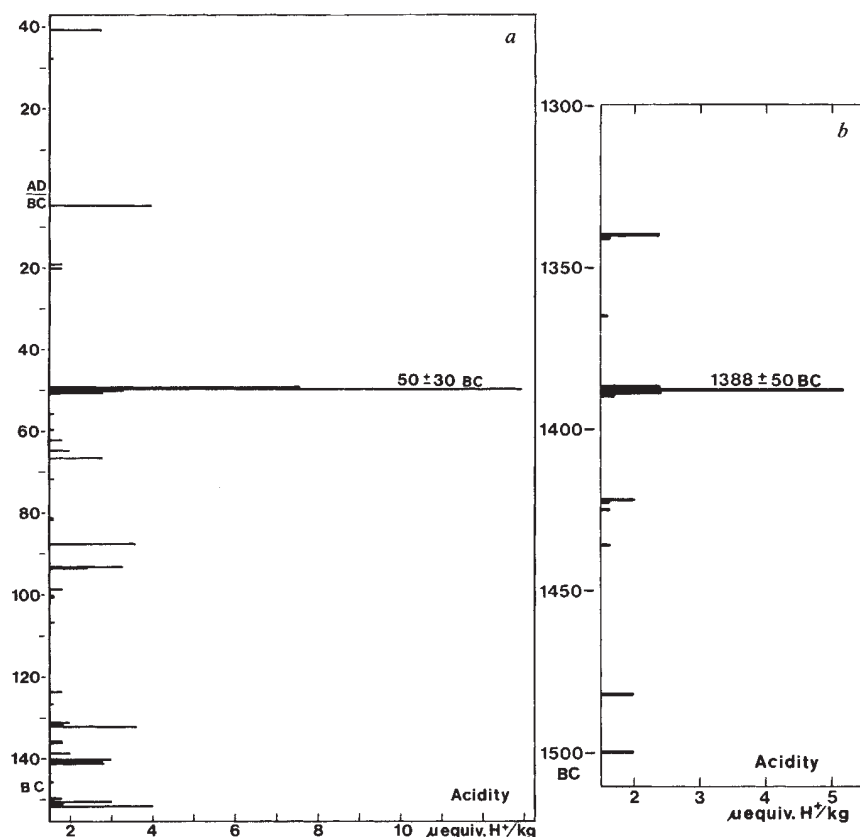


Fig. 4 Unambiguous volcanic signals in Camp Century ice core increments spanning the AD 43–140 BC, and the 1300–1500 BC time intervals. *a*, The unusually large event at $50 \pm 30 \text{ BC}$ may be the one that caused the optical phenomena in the atmosphere described at Julius Caesar's death 44 BC. *b*, The only high signal in the entire period (in fact between 1120 and 2690 BC) occurred at $1388 \pm 50 \text{ BC}$. With the only reservation that another high signal might be lost in the few per cent of missing core, the observed event can be identified with the large Thera eruption in the Aegean Sea that deeply influenced the ancient Minoan society.

carbon datings of which range from 4200 to 4600 BC²⁹. The amount of freshly fallen ash is estimated at 30 km^3 .

$5400 \pm 120 \text{ BC}$: Agrees with the Hekla eruption H5, radio-carbon dated at $5450 \pm 190 \text{ BC}$ ²², and with the largest Thjórsá lava flow, dated by Thorarinsson at $\sim 5500 \text{ BC}$.

None of the high signals before 6000 BC can be safely ascribed to any eruption known by us. Not less than seven great eruptions seem to have occurred between 7000 and 8000 BC. According to Thorarinsson (personal communication), at least some of these must be Icelandic shield volcanoes, such as Skjaldbreiður (15 km^3 ; $\sim 10,000 \text{ yr}$ ago).

No unambiguous evidence of volcanic fallout has been found in the deep ice, which was deposited during the last glaciation, judging from very low δ values in the complete δ profile¹⁹. As the volcanic activity was hardly negligible in this period³⁰, our measuring technique, at least in the present form, obviously cannot be applied to the Camp Century core before 8000 BC. One reason is the limited resolution power of the electrodes, but the main reason is that, disregarding the interstadials revealed by intermediate δ values, the ice from the last glaciation is slightly alkaline, which makes the current between the electrodes nearly zero. In alkaline ice the chemical element content measured so far³¹ is higher than in post-glacial time by factors of, for example, 1.5–3 for Na and SO_4^{2-} ; 2.5–11 for K, Mg, Si and Al; and 10–30 for Ca. The elevated sulphate values are probably just part of the generally increased impurity background, and measuring the volcanic contribution calls for very detailed sulphate analyses. The extremely high calcium values probably originate from CaCO_3 dust from the exposed continental shelves³¹, indicating an alkaline aerosol able to neutralize an acid aerosol that would otherwise cause acidities of up to $8 \mu\text{equiv H}^+ \text{ per kg ice}$.

In preglaciation ice, characterized by high δ values, the chemical impurity concentrations and the acidity background come close to post-glacial values. There is some indication of periods of intense volcanism, but the resolution is too poor for further details.

The climatic impact of volcanism

Volcanic aerosols influence the radiation balance in the atmosphere. The optical thickness τ of an atmospheric aerosol is an index of the loss of radiation by scattering and absorption (the optical thickness of a layer is τ , if the fraction $\exp(-\tau)$ of radiation impinging perpendicularly on the layer is transmitted). τ is an important parameter in radiation balance models for calculating the temperature effect of volcanic aerosols in the stratosphere. The optical properties of an assumed sulphuric acid aerosol have been estimated³ using the measured global 'average' optical depths since 1882. It was thereby possible to calculate temperature changes (ΔT) close to those observed in the atmosphere after volcanic eruptions, including the Agung eruption 1963 that caused up to several degrees warming of the low latitude stratosphere³² and a 0.4°C cooling of the low-latitude troposphere³³. However, an important drawback in these calculations is "that the sign and magnitude of ΔT depends critically on the real and imaginary parts of the refractive indices of the added particles, their size distribution, and the surface albedo"³³, which, of course, stresses the need for a direct comparison between volcanic activity and climatic changes before the instrumental epoch.

The Crête excess-acidity data given in Fig. 1 must be expected to be proportional to the volcanic contribution $\Delta\tau$ to the optical depth in the high northern latitude atmosphere. As a first approximation we further assume that this $\Delta\tau$ is just as relevant for comparison with climatic temperature changes as is the global averaged $\Delta\tau$. This is often implicitly assumed in model calculations, where the latitude-dependent back scattering has been averaged.

Figure 5e shows the Crête acidity profile in 50-yr averaged values. Note that periods of low volcanic activity have been marked in black. The maximum deviation from the overall mean is $< 0.5 \mu\text{equiv H}^+ \text{ kg}^{-1}$ (measured on solid ice⁹), corresponding to only 0.03 and 0.05 deviation of pH measured on melted samples, assuming saturation with CO_2 in ambient air at 0° and

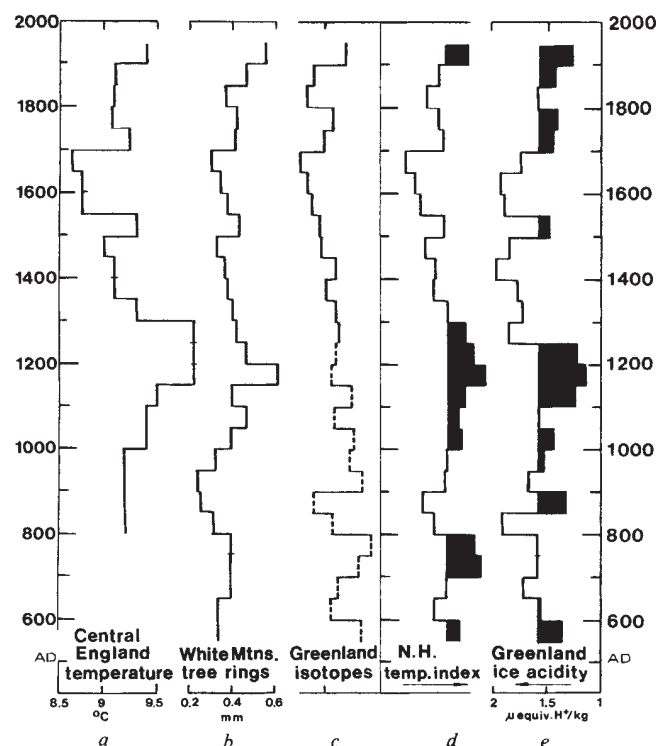


Fig. 5 Correlation between climatic temperature variations (curve *d*, composed of the three indirectly determined records *a* through *c*; warm periods are set out in black) and volcanic activity (curve *e*, showing an acidity profile along the Crête ice core; lower than normal volcanism corresponds to low acidities set out in black).

25 °C, respectively. This might be why 'remarkably stable' 50-yr mean pH values in meltwater from most of the Devon Island ice core³⁴ were interpreted to indicate that 'volcanic activity has not altered significantly' during the period 50–5300 yr BP.

To examine the possible influence of volcanic activity on climate we compare below the Crête acidity record with some absolute-dated climatic records of proxy data reaching far beyond the period of direct temperature observations. As the ice acidity data are more or less representative of aerosol conditions in the Northern Hemisphere in general, they ought to be compared with temperatures of similar general validity, rather than data representative of regional climatic conditions, for example, central England temperature estimates (Fig. 5*a*; ref. 35); or white Mountains, California, tree ring widths (Fig. 5*b*; ref. 36); or the long-term δ -variation in Greenland ice sheet precipitation (Fig. 5*c*; combination of four normalized δ -profiles back to AD 1250; two back to AD 550; ref. 37). But, if

these three climatic curves are combined after normalization, the resulting record (Fig. 5*d*) may be a first approximation to the Northern Hemisphere temperature index needed for making a useful comparison with past volcanic activity. Unfortunately, most other long-term climatic records do not meet the strict requirement to the dating accuracy needed in this context.

The Northern Hemisphere temperature index in Fig. 5*d* does not have a temperature scale, of course. Most of its trends are recognized, more or less, in each of its components—the warm first half of the twentieth century; the cooler nineteenth century; the relatively warm eighteenth century; the 'Little Ice Age' AD 1350–1700 with a minor 'interstadial' AD 1500–1550 (or AD 1350–1900 with 'interstadials' 1500–1550 and 1700–1800), and the mediaeval warmth AD 1000–1300, although the latter seems to have occurred earlier in Greenland.

Many of the above-mentioned features can be recognized in the Crête acidity curve (Fig. 5*e*), warmth generally occurring in periods of lower than average volcanic activity. The correlation coefficient between curves 5*d* and 5*e* is $R = -0.52$ ($P > 95\%$ for 12 d.f.). The curves exhibit a few clear discrepancies, particularly between the data in the AD 850–900 interval. Such occasional discrepancies are not surprising in view of the increasing uncertainty of the temperature index backward in time (for example Lamb gives only one mean temperature estimate for the entire period AD 800–1000). Furthermore, curve *e* is hardly the best obtainable representative of volcanic activity. It could, for example, be improved by replacing the simple acidity measurements by a more differentiated chemical investigation.

The significant correlation between curves *d* and *e* (Fig. 5) indicates that persistent volcanic activity north of 20° S is probably an important, but not necessarily the only, cause of climatic fluctuations of up to several hundred years duration at mid- and high latitudes. If true, a solution of the extremely important and difficult problem of predicting climatic changes by atmospheric model calculations implies a scarcely easier prediction of volcanic activity.

A priori, the acid aerosol load at low latitudes cannot be expected to vary in parallel with curve *e* (Fig. 5) nor, therefore, with the Northern Hemisphere temperature index (curve *d* Fig. 5). This suggests that future climate models should account for the latitudinal variation of the optical depth of volcanic aerosols, rather than using a global average.

A similar treatment of the Camp Century data has been attempted with some success and will be reported elsewhere.

We thank Professor Chester C. Langway, for permission to analyse the Camp Century ice core, and Dr S. Thorarinsson for valuable help. The Danish Commission for Scientific Investigations in Greenland funded the study.

Note added in proof: In a new ice core from South Greenland, a strong acidity signal was found and preliminarily dated at AD 540 ± 10. This might be the Yukon eruption mentioned above. If so, the global acid fall out was of the order of 70 million tons.

Received 4 June; accepted 26 August 1980.

- Lamb, H. H. *Phil. Trans. R. Soc. A* **226**, 425–533 (1970).
- Plinius the Elder *Historia Naturalis* II, 98 (–AD 70).
- Baldwin, B., Summers, A., Toon, O. B., Sagan, C. & Van Camp, W. *Nature* **263**, 551–555 (1976).
- Hansen, J. E., Wang, W.-C. & Lacis, A. A. *Science* **199**, 1065–1068 (1978).
- Mossop, S. C. *Nature* **203**, 824–827 (1964).
- Pollack, J. B. *et al. J. geophys. Res.* **81**, 1071–1083 (1976).
- Castleman, A. W., Munkelwitz, H. R. & Manowitz *Tellus* **26**, 222–234 (1974).
- Hammer, C. U. *Nature* **270**, 482–486 (1977).
- Hammer, C. U. *J. Glaciol.* **93** (in the press).
- Hammer, C. U. *et al. J. Glaciol.* **20**, 3–26 (1978).
- Gudmandsen, P. & Overgaard, S. *Rep. P. 312* (Electromagnetic Institute, Lyngby, 1978).
- Stommel, H. & Stommel, E. *Scient. Am.* **240**(6), 134–140 (1979).
- Larsen, G. *Náttúrfredningurinn* (Reykjavík) **49**, 1–26 (1979).
- Johnsen, S. J. *Proc. I.U.G.G. Symp. Isotopes and Impurities in Snow and Ice*, Grenoble, 210–219 (1975).
- Grasshoff, K. *Methods of Sea Water Analysis* 137–145 (Chemie, Weinheim, 1976).
- Rood, R. T., Sarazin, C. L., Zeller, E. J. & Parker, B. C. *Nature* **282**, 701 (1979).
- Risbo, T., Clausen, H. B. & Lund-Rasmussen, K. (in preparation).
- Macdonald, G. A. *Volcanoes* (Prentice-Hall, New Jersey, 1972).

- Johnsen, S. J., Dansgaard, W., Clausen, H. B. & Langway, C. C. *Nature* **235**, 429–434 (1972).
- Sapper, K. *Katalog der geschichtlichen Vulkanausbrüche* (Schriften, Strassburg, 27 Heft, 1917).
- Virgil *Georgicon* I, 466 (37–31 BC).
- Tauber, H. *Am. J. Sci., Radiocarbon Suppl.* **2**, 5–11 (1960).
- Larsen, G. & Thorarinsson, S. *Jökull* **27**, 28 (1977).
- Watkins, N. D. *et al. Nature* **271**, 122–126 (1980).
- Olsson, I. A. & Broecker, W. S. *Am. J. Sci. Radiocarbon Suppl.* **1**, 1 (1959).
- Popham, M. *Antiquity* **53**, 57–59 (1979).
- Brun, M., Levin, I., Münnich, K. O. & Hubberten, H. H. *Proc. 10th inter. Radiocarbon Conf.*, Bern; *Radiocarbon* (in the press).
- Marinatos, S. *Antiquity* **13**, 425–439 (1939).
- Rubin, M. & Alexander, C. *Am. J. Sci. Radiocarbon Suppl.* **2**, 129–185 (1960).
- Bryson, R. A. & Goodman, B. M. *Science* **207**, 1041–1044 (1980).
- Cragin, J. H., Herron, M. M., Langway, C. C. & Klouda, G. *SCOR/SCAR Polar Ocean Conf.*, Montreal (1974).
- Newell, R. E. *J. Atmos. Sci.* **27**, 977–978 (1970).
- Newell, R. E. & Weare, B. C. *Science* **194**, 1413–1414 (1976).
- Fisher, D. A. & Koerner, R. M. *Geoblastics* (in the press).
- Lamb, H. H. *The Changing Climate* 186 (Methuen, London, 1968).
- LaMarche, V. C. *Jr Science* **183**, 1043–1048 (1974).
- Hammer, C. U., Clausen, H. B. & Dansgaard, W. *Proc. I.U.G.G. Symp. on Volcanism and Climate*, Canberra (in the press).

Nucleotide sequence of cDNA coding for Semliki Forest virus membrane glycoproteins

Henrik Garoff, Anne-Marie Frischaut, Kai Simons, Hans Lehrach & Hajo Delius

European Molecular Biology Laboratory, Postfach 102209, 6900 Heidelberg, FRG

The genes coding for the three membrane polypeptides of Semliki Forest virus have been sequenced and the primary structures of the proteins deduced. The amino acid sequence gives further insight into how the transmembrane structure of the three-chain virus membrane glycoprotein is generated in the infected cell.

MANY enveloped animal viruses acquire their membranes at the host cell surface¹. The virus membranes are highly differentiated plasma membranes which can be studied as models for plasma membrane assembly²⁻⁷. Semliki Forest virus (SFV) consists of a nucleocapsid surrounded by a lipid envelope. The virus membrane is studded with glycoprotein spikes which are essential for virus attachment and penetration into host cells⁸. Each spike glycoprotein contains three different glycopolypeptides: E1 (molecular weight (M_r) 49,000), E2 (M_r 52,000) and E3 (M_r 10,000)^{9,10}. E3 is external to the lipid bilayer whereas E1 and E2 are attached by short hydrophobic polypeptide segments to the membrane. E2 spans the membrane, having about 30 amino acid residues of its carboxy terminus internal to the bilayer¹¹. In infected cells, two species of virus mRNA are found—the 42S RNA, the virus genome and the 26S RNA¹². The latter RNA molecule is homologous to the 3' end of the 42S RNA¹³ and serves as mRNA not only for the E1, E2 and E3 polypeptides but also for the capsid protein of the virus nucleocapsid. All four polypeptides are translated from the 26S RNA using a single initiation site in the order: capsid, p62 and E1^{14,15}. The p62 polypeptide is the precursor of E3 and E2⁹. The capsid protein is cleaved from the nascent chain immediately after it is completed and is left in the cytoplasm¹⁶. The amino-terminal end of p62 then directs the polysomes to the endoplasmic reticulum where the synthesis of the glycoproteins continues concomitantly with membrane translocation and glycosylation¹⁶. The newly synthesized p62 and E1 glycopolypeptides form a complex which migrates through the intracellular membrane system to the plasma membrane, where the p62 protein is cleaved to form E2 and E3¹⁷.

We now report the nucleotide sequence of the cloned cDNA made from the 26S RNA, which codes for the SFV membrane proteins. The amino acid sequence suggests how the glycoproteins are attached to the membrane and provides further insight into how one mRNA directs the synthesis of four polypeptides destined for two different locations of the cell.

Cloning of membrane protein genes

The viral genome (42S RNA) obtained from virus particles and 26S RNA isolated from infected cells were used as templates for cDNA synthesis from the 3' end of the RNA molecules¹⁸. The first strand was made with reverse transcriptase and the second with DNA polymerase¹⁹. The double-stranded cDNA was inserted into the vector pBR322 at the *Pst*I site using the oligo(dG-dC) tailing procedure described by Rowekamp and Firtel²⁰. The hybrid plasmid was then used to transform *Escherichia coli* χ 1776 (ref. 21). Tetracycline-resistant colonies, which hybridized strongly to ³²P-labelled cDNA made from 42S RNA²², were screened to determine the length of the insert molecule²³.

Several clones containing hybrid plasmids with inserts larger than 1 kilobase were found. Electron microscopic analysis of hybrids^{18,24,25} formed between the 42S RNA and DNA inserts showed that three hybrid plasmid clones—pBR SFV 1, 2 and

3—covered the entire 26S RNA region with the exception of about 150 bases from the 3' end. The insert molecule in pBR SFV 1 contained about 2.0 kilobases from the 3' end of the 26S RNA. pBR SFV 2 contained some 2.25 kilobases from the middle region of the 26S RNA. It overlapped by about 1.35 kilobases with the insert of pBR SFV 1. The insert of pBR SFV 3, made from the 42S RNA, contained 1.6 kilobases and was some 250 bases longer than the 5' end of the 26S RNA. It had an overlap of 300 bases with the pBR SFV 2 insert. All inserts were mapped with restriction endonucleases using the technique described by Smith and Birnstiel²⁶. Figure 1 shows agarose gel analyses of the partial digestion products of pBR SFV 1 and pBR SFV 2. The restriction endonuclease mapping of pBR SFV 3, which covers the entire gene of the capsid protein, is presented elsewhere¹⁸. Figure 2 shows the restriction map of the region coding for the membrane proteins. The capsid protein gene ends some 100 bases away from the *Xho*I site in the direction towards the 5' end¹⁸.

Nucleotide sequencing

DNA fragments for sequencing were prepared using two different approaches. The first one was based on the restriction map (Fig. 2). The DNA was cleaved using a suitable restriction endonuclease. After labelling either the 3' or the 5' ends of the DNA fragments, these were cleaved a second time to produce pieces of DNA that had only one labelled end²⁷. The various fragments were isolated and sequenced as described in Fig. 2 legend. In the second procedure we deleted *in vitro* the DNA region of the vector between the *Eco*RI site and the *Pst*I site (748 base pairs) together with different-sized portions of the insert molecule (see Fig. 2 legend). When these molecules were re-cloned, it was possible to sequence directly from the *Eco*RI site of the subcloned plasmid into the insert region (see Fig. 2 legend). This sequencing procedure will be presented in detail elsewhere (A.-M. F., H.G. and H.L., in preparation). The sequenced stretches of DNA are indicated by the arrows in Fig. 2. The sequences were overlapped using a computer program (H. L., unpublished). Figure 3 shows the nucleotide sequence of the region coding for the membrane proteins. The C at the 3' end of the sequence corresponds to the nucleotide before the poly(A) tail in the 26S RNA³⁰. We have included the nucleotide sequence on the 5' side of the *Xho*I site that codes for the carboxy terminus of the capsid protein¹⁸. The complete sequence of the capsid protein will be published elsewhere¹⁸.

Amino acid sequence

Translation of the nucleotide sequence into amino acids using the same reading frame that was earlier used to deduce the primary structure of the capsid protein resulted in a continuous sequence of amino acids (Fig. 3). Reading in the two other frames was interrupted by 40 and 45 nonsense codons, respectively. The coding region of the 26S RNA thus starts with the amino terminus of the capsid protein and terminates 3,759

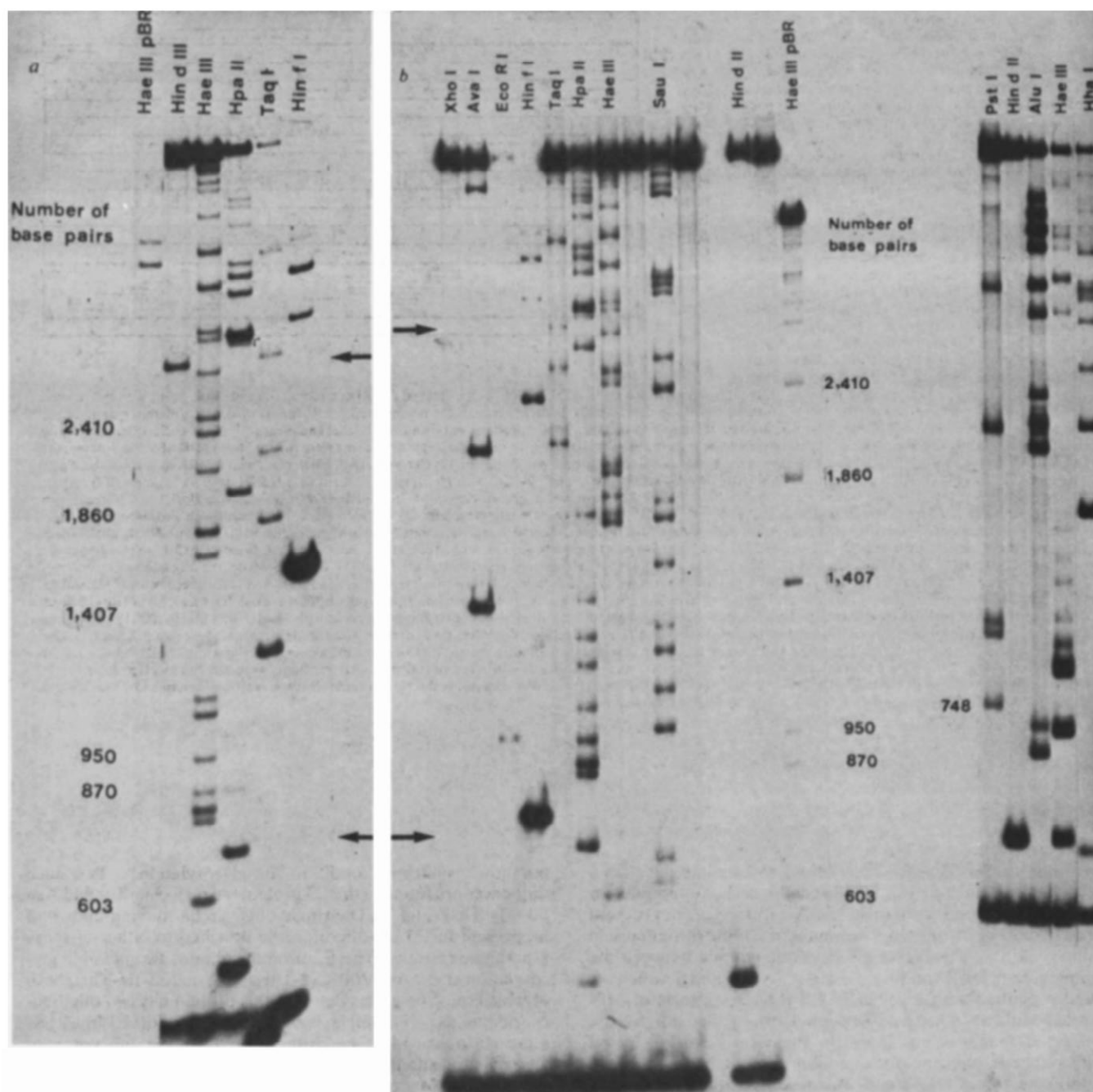


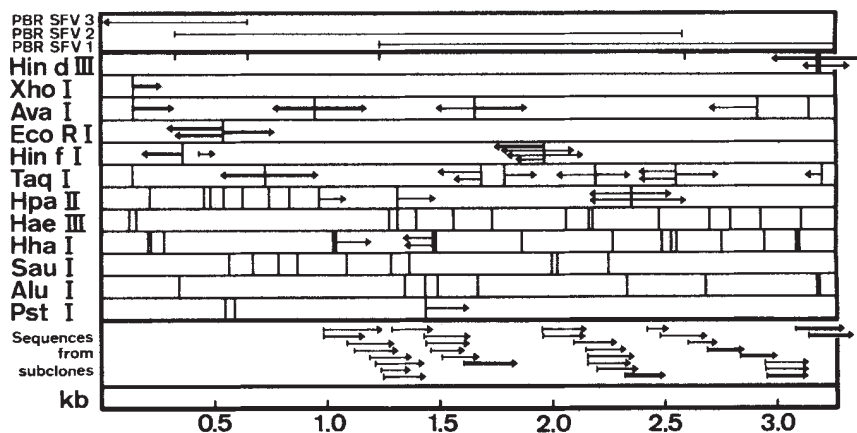
Fig. 1 Analyses of partial digestion products of end-labelled pBR SFV 1 (*a*) and pBR SFV 2 (*b*) by electrophoresis on 1.4% agarose gels (20 cm × 40 cm, 1.5-mm thick). pBR SFV 1 (10 µg) was linearized with *Eco*RI and the same amount of pBR SFV 2 was cleaved with *Cla*I. The 5' ends were dephosphorylated and labelled using polynucleotide kinase and 32 P-ATP (ref. 27). After a second cleavage with *Bam*HI, both DNA samples were subjected to partial digestions with the restriction endonucleases indicated at the top of the figure. The exact incubation conditions as well as the electrophoresis procedure, the drying of the gel and the autoradiography are described elsewhere¹⁸. Fragments generated by cleavages within the insert of pBR SFV 1 are distributed between the two arrows in *a*. Cleavage sites in the insert molecule of pBR SFV 2 are shown by the bands between the arrows in *b*. Some of the digestions of pBR SFV 2 were either too complete or incomplete. These are repeated in a second experiment (right part of *b*). Here the smallest (748 base pairs) and the largest *Pst*I fragments indicate the limits of the fragments derived from the insert. The partial digestion products of pBR322 that were end labelled in the *Eco*RI site were used as standards²⁸.

nucleotides later when the first stop codon (TAA in Fig. 3) is reached. The regions coding for the E3, E2 and E1 polypeptides were localized in this sequence using data available from amino-terminal amino acid sequencing (refs 31, 32 and H. Garoff, H. Riedel and H. Lehrach, in preparation) and carboxypeptidase digestions of the isolated polypeptides³³.

The amino terminus of the isolated E3 polypeptide contains a methionine residue in position 8 and leucine residues in positions 4 and 11 (W. Welch and B. Sefton, personal communication). This amino acid sequence is found immediately after the region coding for the capsid protein (Fig. 3). Bonatti

and Blobel have reported a highly homologous sequence at the amino terminus of the precursor for the E3 and E2 polypeptides from the closely related Sindbis virus³². The carboxy-terminal sequence His-Arg, determined from the isolated E3 polypeptide³³, is found 64 amino acid residues from the carboxy terminus of the capsid protein preceding the known amino terminus of E2 (Ser-Val-Ser-Gln-His-Phe)³¹. As there are two arginine residues before the amino terminus of E2, the last one might be excised by a host protease during cleavage of the p62 polypeptide to E3 and E2 (see below). The amino-terminal sequence determined from the isolated E1 polypeptide

Fig. 2 Restriction endonuclease map of the SFV membrane protein region in pBR SFV 1-3. Thick arrows indicate stretches of DNA where both strands have been sequenced. Thin arrows indicate either 3' end or 5' end sequences. There are no sequence data that overlap the *XhoI* site. However, the corresponding region in the E3 protein is covered with partial amino acid sequence data from a cyanogen bromide peptide (H. Garoff, H. Riedel and H. Lehrach, in preparation). The arrows inside the restriction map represent sequences obtained from isolated DNA fragments and those below the map represent sequences obtained using the subcloning approach. The preparation of subcloned plasmids for sequencing from pBR SFV 1 and 2 will be described elsewhere (A.M.F., H.G. and H.L., in preparation). The steps were as follows: (1) Random linearization of the hybrid plasmid with DNase I. (2) Ligation of *EcoRI* linkers to the ends. (3) Cleavage of the linkers and the vector with *EcoRI*. (4) Separation of DNA molecules by electrophoresis in agarose gels and elution of such molecules which were expected to have deletions extending from the *EcoRI* site on the vector to different positions in the insert. (5) Ligation of the cleaved *EcoRI* octamer at the end of the insert with the cleaved *EcoRI* site of the vector. (6) Transformation and isolation of subcloned plasmids. In these plasmids the regenerated *EcoRI* site is adjacent to different regions of the original insert. The subcloned plasmids were prepared for sequencing using the simple approach: cut with *EcoRI*, label and recut with *ClaI*. As *ClaI* cleaves only 20 bases from the *EcoRI* site, the plasmid preparation could be sequenced without further purification steps. Fragments were isolated from pBR SFV 3 as follows: (1) Cut with *XhoI*, label, recut with *ClaI* which cleaves in the vector, isolate 4,106-base pair fragment. (2) Cut with *AvaI*, label, recut with *EcoRI*, isolate 408-base pair fragment. (3) Cut with *EcoRI*, label, recut with *XhoI*, isolate 408-base pair fragment. (4) Cut with *HinfI*, label, (i) recut with *HaeIII* (vector), isolate 422-base pair fragment, (ii) isolate 697-base pair fragment, recut with *XhoI*, isolate 224-base pair fragment. Fragments from pBR SFV 2: (1) Cut with *TaqI*, label, (i) recut with *HindIII* (vector), isolate 701-base pair fragment, (ii) recut with *HaeIII*, isolate 546-base pair fragment. (2) Cut with *AvaI*, label, recut with *PstI*, isolate 620- and 488-base pair fragments. (3) Cut with *HpaII*, label, isolate 345-base pair fragment, recut with *HaeIII*, isolate 307-base pair fragment. (4) Cut with *HhaI*, label, isolate 437-base pair fragment, recut with *HpaII*, isolate 277- and 160-base pair fragments. (5) Cut with *AvaI*, label, recut with *PstI*, isolate 1,612-base pair fragment. Fragments from pBR SFV 1: (1) Cut with *HpaII*, label, (i) isolate 1,036-base pair fragment, recut with *HinfI*, isolate 648- and 388-base pair fragments, (ii) recut with *HindIII*, isolate 830-base pair fragment. (2) Cut with *PstI*, label, recut with *HinfI*, isolate 521-base pair fragment. (3) Cut with *AvaI*, label, recut with *HinfI*, isolate 1,789- and 950-base pair fragments. (4) Cut with *TaqI*, label, (i) isolate 865-base pair fragment, recut with *HpaII*, isolate 369-base pair fragment, (ii) isolate 405-base pair fragment, recut with *HinfI*, isolate 176- and 229-base pair fragments, (iii) isolate 354-base pair fragment, recut with *HpaII*, isolate 159- and 195-base pair fragments, (iv) isolate 1,768-base pair fragment, either recut with *HindIII*, isolate 635-base pair fragment, or recut with *HaeIII*, isolate 86-base pair fragment. (5) Cut with *HinfI*, label, (i) recut with *HpaII*, isolate 648-base pair fragment, (ii) recut with *HindIII*, isolate 1,218-base pair fragment. (6) Cut with *HindIII*, label, (i) recut with *EcoRI*, isolate 2,708-base pair fragment, (ii) recut with *BamHI* (vector), isolate 3,320-base pair fragment. All fragments were isolated by electrophoresis in agarose gels and subsequent electroelution into hydroxylapatite²⁹. DNA sequencing was done using the base-specific cleavage reactions described by Maxam and Gilbert²⁷. We used five reactions: G, G+A, C+T, C, C+A. kb, Kilobases.



(Tyr-Glu-His-Ser-Thr-Val)³¹ is found 482 amino acids from the amino terminus of E2. However, the carboxy terminus of E2, which has been shown to be His-Ala (ref. 33), is not present immediately before the amino terminus of E1, but 60 amino acid residues earlier. This leaves a 60-residue peptide between the regions coding for E2 and E1. Recently, Welch and Sefton have found a peptide with a M_r of 6,000 (6K) in extracts of cells infected with SFV³⁴. The 6K peptide was coded for by the 26S RNA, but was not found in the virus particle. Partial amino acid sequencing showed that the 6K peptide has a methionine at position 7 and leucines at positions 10 and 17 (W. J. Welch and B. M. Sefton, personal communication), in agreement with the sequence of the 60-amino acid residue segment beginning after the carboxy terminus of E2. The E1 polypeptide terminates at the TAA 438 amino acids after its amino terminus. Carboxypeptidase experiments have shown that the E1 polypeptide terminates with an arginine residue³³, in agreement with the sequence shown in Fig. 3.

The deduced amino acid sequences of the polypeptides have been fully confirmed by partial amino-terminal amino acid sequencing of fragments obtained by cyanogen bromide cleavage of the isolated E1, E2 and E3 polypeptides (H. Garoff, H. Riedel and H. Lehrach, in preparation). The calculated amino acid compositions of the E3, E2 and E1 polypeptides are in good agreement with those obtained experimentally⁹.

Sites of glycosylation

All three membrane polypeptides contain carbohydrate units linked to asparagines^{35,36}. Carbohydrate units in glycoproteins are known to appear either in Asn-X-Ser or in Asn-X-Thr sequences, the presence of these sequences being a necessary

but not a sufficient condition for glycosylation³⁷. Two such sequences are found in the E3 protein (Asn 13-Ala-Thr and Asn 60-Gly-Thr) (Fig. 3). The single oligosaccharide unit known to be present in E3 can therefore be attached to either of these asparagine residues. The E2 protein contains two possible glycosylation sites (Asn 200-Cys-Thr and Asn 262-Ile-Thr), both of which must be glycosylated as this protein has two oligosaccharide units. The single carbohydrate unit of E1 must be attached to Asn 141-Gln-Thr because this is the only potential glycosylation site in the E1 protein.

When the M_s s of the protein (from the sequence) and the carbohydrate parts (from previous gel filtration data) are summed for E1 (47,386+3,400), E2 (46,755+5,100) and E3³⁵ (7,369+4,000), there is good agreement with the apparent M_s s determined for the proteins by SDS gel electrophoresis⁹.

Location of the transmembrane segments

We used the criteria devised by Segrest and Feldman³⁸ to search for hydrophobic sequences in the SFV proteins. Peptides longer than 10 amino acids and lacking charged residues (lysine, arginine, glutamic acid and aspartic acid) were identified and plotted according to their hydrophobicity per unit length. Most of the sequences that were found fall within the triangular area where apolar peptides of globular proteins map. Four peptides mapped in the region to the right of the triangle where the transmembrane domain of glycoporphin³⁹ and the hydrophobic domains of the haemagglutinins of influenza virus^{40,41} are located (Fig. 4). Three of these peptides were in the SFV membrane proteins, the fourth was in the non-structural 6K peptide. Peptide 24 from E1 (Fig. 4) and peptide 10 from E2 are

GAA GAG TGG TCC GCC CCG CTG ATT ACT GCC ATG TGT GTC CTT GCG AAT GCT AGC TTC CCG TGC TTC GAG CCC CCG GLU GLU TRP SER ALA PRO LEU ILE THR ALA MET CYS VAL LEU ALA ASN ALA THR PHE PRO CYS PHE GLN PRO PRO C E3	E3(47)	GCT TAC GAA CAT TCG ACA GTA ATG CCG AAC CTG CTG CCG TTC CCG TAT AAG GCT CAC ATT GAA AGC GCA GCA TAT ALA TYR GLU HIS SER THR VAL MET PRO ASN VAL VAL GLY PHE PRO TYR LYS ALA HIS ILE GLU ARG PRO GLY TYR 6K E1	E1(49)
TCT GTA CCT TCC TGC TAT GAA AAC AAC CCA GAG CCC ACA CTA CCG ATC CTC GAG GAT AAC CTC GAT AGC CCA GCG CYS VAL PRO CYS CYS TYR GLU ASN ASN ALA GLU ALA THR LEU ARG MET LEU GLU ASP ASN VAL ASP ARG PRO GLY	E2(31)	GAC TAC AAC ACC CTC CTC CCG TCG CCG TAC CTG AAC TGC TGC GCG GCG TCA GAG TGC TCC ACT AAA GAG AAC CCT GLU TYR LYS THR VAL VAL PRO SER PRO TYR VAL LYS CYS CYS GLY ALA SER GLU CYS SER THR LYS GLU LYS PRO	E1(74)
TAC TAC GAC CTC CTT CAG CCA CCC TTC ACC TCC CCA AAC GGA ACA AGA CAC CCG CCG ACC GTC TCG CAA CAC TTC TYR TYR ASP LEU LEU GLN ALA ALA LEU THR CYS ARG ASN GLY THR ARG HIS ARG ARG SER VAL SER GLN HIS PHE E3 E2	E2(56)	CAC TAC CAA TCC AAC CTT TAC ACA CCG CTG TAC CCG TTC ATG TCG GCA GCG GCA TAT TCC TTC TCG GAC TCA GAA ASN VAL TYR LYS ALA THR ARG PRO TYR ILE ALA TYR CYS ALA ASP CYS GLY ALA GLY HIS SER CYS HIS SER PRO	E1(99)
GTA GCA ATT GAA CCG CTC AGC TCC GAA GCT ACC GAG CCG ATG CTG AAG ATT CAG TTC TCG CCA CAA ATT GCG ATA VAL ALA ILE GLU ALA VAL ARG SER GLU ALA THR ASP GLY MET LEU LYS ILE GLN PHE SER ALA GLN ILE GLY ILE	E2(81)	AAC ACC CAA CTC ACC GAG CCG TAC CTC GAT CCA TCC GAC CTA TCC AGC CAT CAT CAG CCA TCT OCT TAC AAA CCG ASN THR GLN LEU SER GLU ALA TYR VAL ASP ARG SER ASP VAL CYS ARG HIS ASP HIS ALA SER ALA TYR LYS ALA	E1(124)
GAT AAG AGT CAC AAT CAT CAC TAC ACC AAG ATA AGC TAC CCA GAG GCG CAC GCG ATT CAG AAT GCG CTC CCG TCA ASP LYS SER ASP ASN HIS ASP TYR THR LYS ILE ARG TYR ALA ASP GLY HIS ALA ILE GLU ASN ALA VAL ARG SER	E2(106)	CAT ACA CCA TCG CTC AAC CCG AAA CTC AGC GGT ATC TAC CCG AAC CTA AAC CAG ACT GTG GAT GTT TAC CTC AAC HIS THR ALA SER LEU LYS ALA LYS VAL ARG VAL MET TYR GLY ASN VAL ASN GLN THR VAL ASP VAL TYR VAL ASN	E1(149)
TCT TTC AAG GTA GCC ACC TCC GGA CAC TGT TTC GTC CAT GCG ACA ATG GGA CAT TTC ATA CTG CCA AAG TCC CCA SER LEU LYS VAL ALA THR ARG PRO TYR ILE ALA TYR CYS ALA ASP CYS GLY ALA GLY HIS SER CYS HIS SER PRO	E2(131)	GCA GAC CAT GCG CTC ACC ATA CCG GGT ACT CAG TTC ATA TTC GCG CCG CTC TCA TCG CCG TCG ACC CCG TTC GAC GLY ASP HIS ALA VAL THR ILE GLY GLY THR GLN PHE ILE PHE GLY PRO LEU SER SER ALA TRP THR THR LYS ASP	E1(174)
CCG GGT GAA TTC CTC CAG GTC TCG ATC CAG GAC ACC AGA AAC CCG GTC CCG CCG TCG ACA ATA CAA TAT CAT CAT PRO GLY GLU PHE LEU GLN VAL SER ILE GLN ASP THR ARG ASN ALA VAL ARG ALA CYS ARG ILE GLN TYR HIS HIS	E2(156)	AAC AAC ATA CTC CTC TAC AAA GAG GAA CTC TTC AAT CAG CAG TTC CCG CCG TAC GGA TCT CCG CAA CCA CCG CCG ASN LYS ILE VAL VAL TYR LYS ASP GLU VAL PHE ASN GLN ASP PHE PRO PRO TYR GLY SER GLY GLN PRO GLY ARG	E1(199)
GAC CCT CAA CCC CTC GGT AGA CAA AAA TTT ACA ATT AGA CCA CAC TAT CCA AAA CAG ATC CCT TCC ACC ACT TAT ASP PRO GLN PRO VAL GLY ARG GLU LYS PHE THR ILE ARG PRO HIS TYR GLY LYS GLU ILE PRO CYS THR THR TYR	E2(181)	TTC CCG GAC ATC CAA AGC ACA ACA CTC CAG ACT AAC GAC CTC TAC CCG AAC ACC CCA CTC AAG CTC GCA CCG CCT PHE GLY ASP ILE GLN SER ARG THR VAL GLU SER ASN ASP LEU TYR ALA ASN THR ALA LEU LYS LEU ALA ARG PRO	E1(224)
GAA CAC ACC ACA CCG GAG ACC CTG CAG GAA ATC CAC ATC CAT ATC CCG CCA GAT ACC CCG CAC ACC ACC TTC CTA GLN GLN THR THR ALA GLU THR VAL GLU GLU ILE ASP MET HIS MET PRO PRO ASP THR PRO ASP ARG THR LEU LEU	E2(206)	TCA CCG GGT ATG CTC CAT CTA CCG TAC ACA CAG ACA CCT TCA GCG TTC AAA TAT TCG CTA AAG GAA AAA GCG ACA SER GLN GLN SER GLY ASN VAL LYS ILE THR VAL GLY GLY LYS LYS VAL LYS TYR ASN CYS THR CYS GLY THR GLY	E1(249)
TCA CAC CAA TCT CCG AAT CTA AAC ATC ACA CTC CCA GGA AAC AAG CTC AAA TAC AAC TCG ACC TGT CCA ACC GGA SER GLN GLN SER GLY ASN VAL LYS ILE THR VAL GLY GLY LYS LYS VAL LYS TYR ASN CYS THR CYS GLY THR GLY	E2(231)	CCC CTA AAT ACC AAG CCG CCG TTT CCG TCC CAA ATC AAA ACC AAC CCT CTC AGC GCG ATG AAC TCG CCG CTC GGA ALA LEU ASN THR LYS ALA PRO PHE GLY CYS GLN ILE LYS THR ASN PRO VAL ARG ALA MET ASN CYS ALA VAL GLY	E1(274)
AAC GTT GCG ACT ACT AAT TCG CAC ATG AGC ATC AAG ACC TGT CTA ATA GAG CAG TCG CAC CTC TCA CTC ACC CAC ASN VAL GLY THR THR ASN SER ASP MET THR ILE ASN THR CYS LEU ILE GLU GLN CYS HIS VAL SER VAL THR ASP	E2(256)	AAC ATG CCT CTC TCC ATG AAT TTC OCT CAG ACC GCG TTT ACC CCG ATT CTC CAG CCG CCG ACC ATC ATT GAC CTC ASN ILE PRO VAL SER MET ASN LEU PRO ASP SER ALA PHE THR ARG ILE VAL GLU ALA PRO THR ILE ILE ASP LEU	E1(299)
CAT AAG AAA TCG CAG TTC AAC TCA CCT TTC GTC CCG AGA GCG CAC GAA CCG CCG TCA AGA AAA GCG AAA CTC CAT ATC HIS LYS LYS TRP GLN THR THR ASN SER PRO PHE VAL PRO ARG ALA CAG GLU PRO THR VAL ILE HIS LYS VAL HIS ILE	E2(281)	ACT TCG ACA CTG GCT ACC TGT ACC CAC TCC TCG CAT TTC GCG GCG CTC TTC ACA CTC ACC TAC AAC ACC AAC AAG THR CYS THR VAL THR ILE GLY GLY THR THR THR THR THR THR THR THR THR THR THR THR THR THR THR THR THR THR	E1(324)
CCA TTC CCG TTC GAC AAC ATC ACA TCG AGA GTT CCA ATC CCG CCG GAA CCA ACC CTC ATC CAC GCG AAA AGA GAA PRO PHE PRO LEU ASP ASN ILE THR CYS ARG VAL PRO MET ALA ARG GLU PRO THR VAL ILE HIS GLY LYS ARG GLU	E2(306)	AAC GCG GAC TCG TCT GTA CAC TCG CAC TCT AAC GTA CCG ACT CTA CAG GAG GCG ACA GCA AAA GTC AAG ACA GCA ASN GLY ASP CYS SER VAL HIS SER HIS SER ASN VAL ALA THR LEU GLN GLU ALA THR ALA LYS VAL LYS THR ALA	E1(349)
CTG ACA CTC CAC CTT CAC CCA GAT CAT CCC ACC CTC TTT TCC TAC CCG ACA CTC GCT CAG CAC CCG CAC TAT CAC VAL THR LEU HIS LEU HIS PRO ASP HIS PRO THR LEU PHE SER TYR ARG THR LEU GLY GLU ASP PRO GLN TYR HIS	E2(331)	GCT AAG CTC ACC TTA CAC TTC TCG ACC CCA ACC GCA TCA CCT TCT TTT CTC CTC TCG CTA TCG ACT CCG ACC GCG GLY LYS VAL THR LEU HIS PHE SER THR ALA SER ALA SER PRO SER PHE VAL VAL SER LEU CYS SER ALA ARG ALA	E1(374)
GAG GAA TCG CTC ACA CCG CCG CTC GAA CCG ACC ATA CCG CTA CCA GTC CAG CCG ATG GAG TAC CAC TCG CCA AAC GLU GLU TRP VAL THR ALA ALA VAL GLU ARG THR ILE PRO VAL PRO VAL ASP GLY MET GLU TYR HIS TRP GLY ASN	E2(356)	ACC TGT TCA CCG TCG TGT GAG CCG CCG AAA GAC CAC ATA CTC CCA TAT CCG GCT ACC CAC ACT AAC CTA CTC TTT THR CYS SER ALA SER CYS GLU PRO PRO LYS ASP HIS ILE VAL PRO TYR ALA ALA SER HIS SER ASN VAL VAL PHE	E1(399)
AAC GAC CCA CTC AGC CTT TCC TGT CAA CTC ACC ACT CAA GCG AAA CCG CAC GCG TCG CCG CAT CAC ATC CTA CAC ASN ASP PRO VAL ARG LEU TRP SER GLN LEU THR THR GLU GLY LYS PRO HIS GLY TRP PRO HIS GLN ILE VAL GLN	E2(381)	CCA GAC ATG TCG CCG ACC CCA CTA TCA TCG CTG CAG AAA ATC TCG GGT GGT CTC GCG CCG TTC GCA ATC CCG CCG PRO ASP MET SER GLY THR ALA LEU SER TRP VAL GLN LYS ILE SER GLY GLY LEU GLY ALA PHE ALA ILE GLY ALA	E1(424)
TAC TAC TAT CCG CTT TAC CCG CCG GCT ACA CTA TCC CCG GTC GTC GCG ATC AGC TTA CTG GCG TTC ATA TCG ATC TYR TYR TYR LEU TYR PRO ALA ALA THR VAL ALA VAL VAL GLY MET SER LEU VAL GLY MET SER LEU VAL GLY MET SER LEU	E2(406)	ATC CTC CTC CTC GTT GTC CTC ACT TCG ATT CCG CTC CCG ACA TAA CTT AGC GTA CCG AAT CCG ATT GAT ATA GCA ILE LEU VAL LEU VAL VAL VAL THR CYS ILE GLY LEU ARG ARG	
TTC CCG TCG TCC TAC ATC CTC GTT CCG CCG CCG AGT AAG TCG TTC ACC GCT TAT GCT TTA ACA CCA GGA CCG CCA PHE ALA SER CYS TYR MET LEU VAL ALA ALA ARG SER LYS CYS LEU THR PRO TYR ALA LEU THR PRO GLY ALA ALA	6K(34)	AGA AAA TTC AAA ACA GAA AAA GTT AGC CTA AGC AAT GCG ATA TAA CCA TAA CTC TAT AAC TTC TAA CAA ACC CCA	
GTT CCG TCG ACC CTC CCG ATA CTC TCG TCC CCG CCG CCG CAC CCA GCT AGT CTC GCA CAG ACT ATC CCG TAC VAL PRO TRP THR LEU GLY ILE LEU CYS CYS ALA PRO ARG ALA HIS ALA ALA SER VAL ALA GLU THR MET ALA TYR E2 6K	6K(59)	ACA AGA CCT CCG CAA TTC CCG CCG TCG TCC CCG TCA CCG AAA CTC GCG CCA ACT CAT ATT GAC ACA TTA ATT GCG	
TTC TCG GAC CAA AAC CAA GCG TTC TTC TCG TTC GAG TTT CCG CCG CCT GTT CCG TCG ATC CTC ATC ACC TAT LEU TRP ASP GLN ALA LEU PHE TRP LEU GLU ALA CYS ALA ILE ALA CYS ALA ILE LEU ILE ILE THR TYR		AAT AAT TCG AAG CTT ACA TAA GGT TAA TTC GAG GAA TAA TTC CAT TTT TAT TTT ATT TTC CAA TTC GTT TTT AAT	
TCC CTC AGA AAC CTC CTC TGT TCG TGT AAG ACC CTT TGT TTT TTA GTC CTA CTC ACC CTC CCG GCA ACC CCG AGA CYS LEU ARG ASN VAL LEU CYS CYS CYS LYS SER LEU SER PHE LEU VAL LEU LEU SER LEU GLY ALA THR ALA ARG		ATT TCG	

Fig. 3 The nucleotide sequence of the membrane protein genes. The deduced amino acid sequences of the membrane proteins are shown below. The amino acid residues are numbered from the amino terminus of each polypeptide. The positions are shown in the parentheses to the right. Membrane-spanning segments are underlined and potential glycosylation sites are marked (●).

long enough to span the lipid bilayer in the form of an α -helix³⁸. These peptides are located in the carboxy-terminal regions of E1 and E2, respectively, where the hydrophobic domains of E1 and E2 have previously been mapped¹¹. The presumptive transmembrane domain of E2 is located between Lys 346 and Arg 392, 31 amino acids away from the carboxy terminus, in complete agreement with our previous biochemical data¹¹. The transmembrane domain of E2 is rather long, as defined in the diagram shown in Fig. 4. However, scrutiny of the amino acid sequence shows that it contains two histidines at positions 348 and 352 and two glutamines at positions 353 and 356. These may be located in the external domain outside the bilayer because these amino acids have not been found in the other known transmembrane domains. Furthermore, if we assume an α -helix for the membrane-spanning segment, the proline residue at position 363 would also be excluded from the bilayer, leaving 28 residues in the transmembrane domain. However, we cannot exclude the possibility that more of the segment is in the bilayer.

The hydrophobic domain of E1 is 24 residues long, only two amino acid residues away from the carboxy terminus. There is no evidence from previous studies that this domain spans the membrane, but the methods we previously used to demonstrate spanning would not have detected the very short segment of two arginine residues^{11,42}. Examination of the known and presumptive transmembrane domains of cell surface^{39,43,44} and virus glycoproteins^{40,41} (including E2) show that they are all characterized by hydrophobic segments at least 23 amino acid residues long (lacking Lys, Arg, Asp, Glu, Gln, Asn, His and Pro residues) followed by one or several closely spaced basic amino acids on their carboxy-terminal side marking the start of the cytoplasmic domain. The hydrophobic domain of the E1 polypeptide fulfills these criteria. Its cytoplasmic domain would thus be only two arginine residues long.

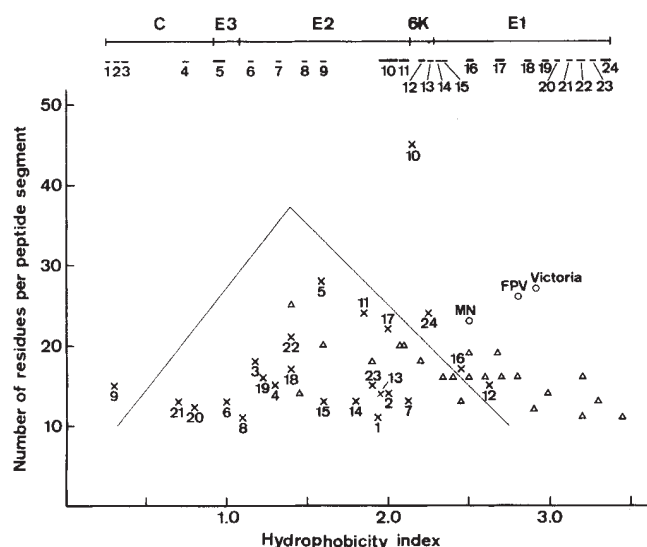


Fig. 4 Hydrophobicity diagram of apolar sequences in SFV proteins. The hydrophobicity of apolar peptides containing 10 or more amino acids was calculated as described by Segrest and Feldman³⁸ and these values were plotted against the lengths of the peptides. The upper part of the figure shows the location of the apolar sequences on the polypeptides translated from the 26S RNA. Symbols: $\times$, apolar peptides from SFV proteins; these are numbered according to their location. $\circ$, Membrane-spanning peptides from glycoporphin (marked MN)³⁹, the haemagglutinin protein from fowl plague virus (FPV)⁴⁰ and influenza virus A strain Victoria⁴¹. Δ , Apolar segments of amino-terminal extensions (signal peptides) of various secreted proteins (see ref. 60).

The third hydrophobic peptide mapping to the right of the triangle in Fig. 4 is located between Lys 79 and Asp 97 in the external hydrophilic domain of E1. This region is amino terminal to the presumptive attachment site for the oligosaccharide side chain. It is 16 amino acids long and is interrupted by a proline residue, much in the same way as the hydrophobic domain of cytochrome *b*₅ (ref. 45). This region could be involved in the membrane fusion activity displayed by the SFV spike glycoprotein during virus entry into the host cell through the lysosomes⁴⁶. At $\leq$ pH 6, the membrane of SFV fuses with cholesterol-containing phospholipid bilayers⁴⁷. One could envisage that the low pH induces a conformational change in the E1 polypeptide, exposing this hydrophobic peptide which could then insert into a neighbouring lipid bilayer to initiate fusion of the apposed membranes. The amino-terminal region of the Sendai virus fusion protein, thought to be involved in membrane fusion, is also relatively hydrophobic⁴⁸.

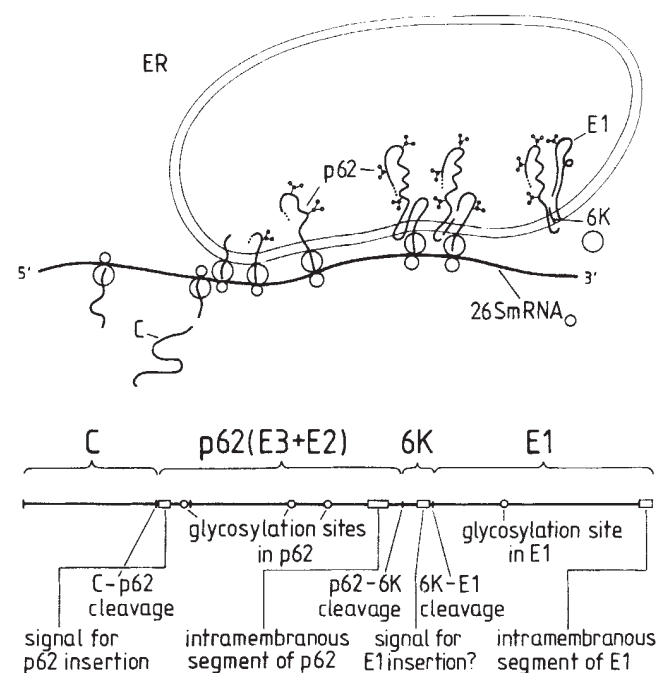


Fig. 5 Scheme for the assembly of SFV membrane glycoprotein into the membrane of the endoplasmic reticulum. C is the capsid protein which remains in the cytoplasm.

Proteolytic cleavages involved in synthesis of SFV proteins

The nucleotide sequence of the DNA, copied from the 26S RNA, has confirmed the order determined previously for the genes coding for the SFV proteins^{11,14}. The coding regions are contiguous except between E2 and E1, where an additional gene segment was found coding for 60 amino acids, the 6K peptide described by Welch and Sefton³⁴. This would imply that at least four proteolytic cleavages are needed to generate the four structural polypeptides of the virus. Previous studies have shown that the capsid protein is cleaved from the nascent chain as soon as translation of the protein is completed¹⁶. This cleavage is mediated either by the capsid protein itself⁴⁹ or by a ribosomal protease⁵⁰, as it has been observed *in vitro* in all eukaryotic protein-translation systems tested. The excision of the 6K peptide also occurs on the nascent chain and may involve host proteases. The fourth cleavage which generates the E3 and E2 polypeptides from the p62 protein occurs about 30 min after synthesis¹⁷. The cleavage is blocked by antibodies against the SFV spike glycoproteins added to the infected cells, and prob-

ably occurs at the plasma membrane^{17,51}. The cleavage site is next to an arginine pair located before the amino terminus of E1. Many prohormones and proproteins are processed to their final forms through specific cleavages at sites bearing pairs of basic amino acid residues (lysine and arginine). This is the case for proinsulin⁵², progastrin⁵³, proglucagon⁵⁴, proalbumin⁵⁵, parathyroid hormone⁵⁶ and the common precursor for melanotropins, corticotropin, lipotropin and endorphins⁵⁷. These cleavages which probably involve a combined trypsin-like and carboxypeptidase B-like specificity occur late in secretion, shortly before extracellular release. The haemagglutinin of fowl plague virus is also split late during intracellular transport or at the plasma membrane by a host protease into two polypeptide chains by two cleavages involving the excision of a -Lys-Arg-Glu-Lys-Arg- connecting peptide⁴⁰. It is tempting to speculate that the secreted proproteins and the virus membrane glycoproteins are processed by the same family of host proteases, part of the post-Golgi pathway to the cell surface. This would mean that the cleavage of the p62 polypeptides begins before their appearance on the cell surface.

Assembly of SFV membrane proteins

Previous studies indicate that the p62 and the E1 polypeptides are transferred across the membrane of the endoplasmic reticulum as they are being synthesized^{16,32}. Both polypeptide chains seem to have their own independent signal peptides for membrane insertion^{16,58,59}. Membrane insertion and transfer of the p62 polypeptide occur *in vitro* only if microsomal vesicles are present in the translation assay before about 100 amino acids from the amino terminus of the p62 polypeptide have been made¹⁶. This region must therefore contain a functional signal peptide. There seems to be no significant homology in the primary structures of the different signal sequences so far known^{7,60}. The only common feature is a central core of at least nine predominantly hydrophobic amino acids uninterrupted by charged amino acids⁶¹. This criterion is fulfilled by the amino-terminal region of the p62 protein (Fig. 3). The hydrophobicity of this peptide (no. 5 in Fig. 4) is low but not lower than that of other signal peptides (Fig. 4). An unusual observation is that the signal peptide of the p62 polypeptide is not removed during translation (ref. 32 and W. Welch and B. Sefton, personal communication). It is completely translocated through the membrane with the rest of the external p62 domain of the protein.

The evidence suggesting that the E1 polypeptide has its own independent signal peptide is based on observations using a temperature-sensitive mutant of SFV. Kääriäinen and co-workers have found a mutant of SFV in which the cleavage between the capsid and the p62 polypeptides is blocked at the nonpermissive temperature⁵⁹. Although the p62 polypeptide is not inserted into the membrane of the endoplasmic reticulum in these conditions, the E1 polypeptide seems to be correctly translocated. Judging from the M_r s of the E1 polypeptide and the uncleaved capsid-p62 polypeptide, the 6K peptide seems also to be normally excised. It is not possible to define the exact location of the signal peptide for E1 from the amino acid sequence of this region (Figs 3, 4), but one possibility is that it is in the 6K peptide preceding the amino terminus of E1. In this case, polypeptide E1 would be released from the 6K peptide, presumably by a signal peptidase cleavage.

A scheme for the assembly of the SFV spike glycoprotein into the membrane of the endoplasmic reticulum is depicted in Fig. 5. The signal peptide of the p62 polypeptide mediates insertion into the membrane soon after capsid protein cleavage. Translocation of the growing chain continues until the carboxy-terminal transmembrane domain stops transfer of the polypeptide. The ribosome would either be detached from the membrane at this stage or remain bound to the membrane. When the signal peptide for E1 emerges from the ribosome, renewed membrane insertion and transfer would take place. We do not know exactly when and on what side of the membrane the cleavage releasing the 6K peptide occurs during translation. It would be helpful to know the location of the 6K peptide with respect to the membrane after excision. Translocation of the E1 polypeptide is not complete when the nonsense codon in the 26S RNA terminates translation but the transmembrane domain of E1 must be pulled into the membrane either by the spontaneous folding of the external glycosylated domain in the lumen of the endoplasmic reticulum or by the postulated translocation machinery in the membrane⁷. The cluster of basic amino residues on the carboxy-terminal side of the membrane-spanning segments may be needed to fix the polypeptides into the membrane.

We thank Evelyn Kiko and Hilka Virta for technical assistance, Annie Biaias for typing the manuscript, N. Kalkkinen, L. Kääriäinen, B. Sefton, J. Strauss for permission to use unpublished data, and A. Helenius, K. Matlin and G. Warren for helpful criticism. This work has been done in accordance with the German guidelines for recombinant DNA research.

Received 7 July; accepted 22 September 1980.

- Simons, K. & Garoff, H. *J. gen. Virol.* (in the press).
- Rothman, J. E. & Lodish, H. F. *Nature* **269**, 775-780 (1977).
- Lingappa, V. R., Katz, F. N., Lodish, H. F. & Blobel, G. *J. biol. Chem.* **253**, 8667-8670 (1978).
- Dobberstein, B., Garoff, H. & Warren, G. *Cell* **17**, 759-769 (1979).
- Ploegh, H. L., Cannon, L. E. & Strominger, J. L. *Proc. natn. Acad. Sci. U.S.A.* **76**, 2273-2277 (1979).
- Jokinen, M., Gahmberg, C. G. & Anderson, L. C. *Nature* **279**, 604-607 (1979).
- Blobel, G. *et al. Proc. Soc. exp. Biol.* **33**, 9-36 (1979).
- Utermann, G. & Simons, K. *J. molec. Biol.* **85**, 569-587 (1974).
- Garoff, H., Simons, K. & Renkonen, O. *Virology* **61**, 493-504 (1974).
- Ziemiacki, A. & Garoff, H. *J. molec. Biol.* **122**, 259-269 (1978).
- Garoff, H. & Söderlund, H. *J. molec. Biol.* **124**, 535-549 (1978).
- Simmons, D. T. & Strauss, J. H. *J. molec. Biol.* **71**, 599-613 (1972).
- Kennedy, S. I. T. *J. molec. Biol.* **108**, 491-511 (1976).
- Clegg, J. C. S. *Nature new Biol.* **254**, 454-455 (1975).
- Glanville, N., Ranki, M., Morser, J., Kääriäinen, L. & Smith, A. E. *Proc. natn. Acad. Sci. U.S.A.* **73**, 3059-3063 (1976).
- Garoff, H., Simons, K. & Dobberstein, B. *J. molec. Biol.* **124**, 587-600 (1978).
- Ziemiacki, A., Garoff, H. & Simons, K. *J. gen. Virol.* (in the press).
- Garoff, H., Frischauf, A.-M., Simons, K., Lehrach, H. & Delius, H. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
- Lehrach, H. *et al. Biochemistry* **18**, 3146-3152 (1979).
- Röwekamp, W. & Firtel, R. *Dev. Biol.* (in the press).
- Villa-Komaroff, L. *et al. Proc. natn. Acad. Sci. U.S.A.* **75**, 3727-3731 (1978).
- Grunstein, M. & Hogness, D. S. *Proc. natn. Acad. Sci. U.S.A.* **72**, 3961-3965 (1975).
- Barnes, W. M. *Science* **195**, 393-394 (1977).
- Chow, L. T., Roberts, J. M., Lewis, J. B. & Broker, T. R. *Cell* **11**, 819-836 (1977).
- Priess, H., Koller, B., Hess, B. & Delius, H. *Molec. gen. Genet.* **178**, 27-34 (1980).
- Smith, H. O. & Birnstiel, M. L. *Nucleic Acids Res.* **3**, 2387-2398 (1976).
- Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 499-560 (1980).
- Sutcliffe, J. G. *Nucleic Acids Res.* **5**, 2721 (1978).
- Tabak, H. F. & Flavell, R. A. *Nucleic Acids Res.* **5**, 2321-2332 (1978).
- Ou, J.-H., Strauss, E. G. & Strauss, J. H. *Virology* (in the press).
- Kalkkinen, N., Jörnwall, H., Söderlund, H. & Kääriäinen, L. *Eur. J. Biochem.* **108**, 31-37 (1980).
- Bonatti, S. & Blobel, G. *J. biol. Chem.* **254**, 12261-12264 (1979).
- Kalkkinen, N. *FEBS Lett.* **115**, 163-166 (1980).
- Welch, W. J. & Sefton, B. M. *J. Virol.* **33**, 230-237 (1980).
- Mattila, K., Luukkainen, A. & Renkonen, O. *Biochim. biophys. Acta* **419**, 435-444 (1976).
- Pesonen, M. & Renkonen, O. *Biochim. biophys. Acta* **455**, 510-525 (1976).
- Neuberger, A., Gottschalk, A., Marshall, R. D. & Spiro, R. G. in *The Glycoproteins: their Composition, Structure and Function* (ed. Gottschalk, A.) 450-490 (Elsevier, Amsterdam, 1972).
- Segrest, J. P. & Feldman, R. J. *J. molec. Biol.* **87**, 853-858 (1974).
- Tomita, H. & Marchesi, V. T. *Proc. natn. Acad. Sci. U.S.A.* **72**, 2964-2968 (1975).
- Porter, H. G. *et al. Nature* **282**, 471-477 (1979).
- Min Jou, W. *et al. Cell* **19**, 683-696 (1980).
- Garoff, H. & Simons, K. *Proc. natn. Acad. Sci. U.S.A.* **71**, 3988-3992 (1974).
- Robb, R. J., Terhorst, L. & Strominger, J. L. *J. biol. Chem.* **253**, 5319-5324 (1978).
- Rogers, J. *et al. Cell* **20**, 303-312 (1980).
- Ozols, J. & Gerard, C. *Proc. natn. Acad. Sci. U.S.A.* **74**, 3725-3729 (1977).
- Helenius, A., Kartenbeck, J., Simons, K. & Fries, E. *J. Cell Biol.* **84**, 404-420 (1980).
- White, J. & Helenius, A. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3273-3277 (1980).
- Gething, M. J., White, J. & Waterfield, M. D. *Proc. natn. Acad. Sci. U.S.A.* **76**, 2737-2740 (1978).
- Aliperti, G. & Schlesinger, M. J. *Virology* **90**, 366-369 (1978).
- Langner, J., Wiederanders, S., Ansoorge, P., Bohley, P. & Krischke, H. *Acta biol. med. germ.* **38**, 1527-1538 (1979).
- Jones, K. J., Waite, M. R. F. & Bose, H. R. *J. Virol.* **13**, 809-810 (1974).
- Kemmler, W., Steiner, D. F. & Borg, J. *J. biol. Chem.* **248**, 4544-4551 (1973).
- Gregory, R. A. & Tracy, H. J. *Lancet* **ii**, 797 (1972).
- Tager, H. S. & Steiner, D. F. *Proc. natn. Acad. Sci. U.S.A.* **70**, 2321-2325 (1973).
- Russell, J. H. & Geller, D. M. *J. biol. Chem.* **250**, 3409-3413 (1975).
- Habener, J. F., Chang, H. T. & Potts, J. T. Jr *Biochemistry* **16**, 3910-3917 (1977).
- Nakanishi, S. *et al. Nature* **278**, 423-427 (1979).
- Wirth, D. F., Katz, F., Small, B. & Lodish, H. F. *Cell* **10**, 253-263 (1977).
- Hashimoto, K., Erdei, S., Keränen, S., Saraste, J. & Kääriäinen, L. *J. Virol.* (in the press).
- Austen, B. M. *FEBS Lett.* **103**, 308-313 (1979).
- Emr, S. D., Hedgpeth, J., Clement, J. M., Silhavy, T. J. & Hoffnung, M. *Nature* **285**, 82-85 (1980).

LETTERS

Stellar perturbations of the cometary cloud

Paul R. Weissman

Jet Propulsion Laboratory, Earth and Space Sciences Division, Pasadena, California 91103

The existence of a cloud of $\sim 2 \times 10^{11}$ comets surrounding the Solar System and extending out to interstellar distances has been suggested by Oort¹ based on the observed distribution of inverse semi-major axes of the long-period comets. Simple statistical arguments can be used to update the current understanding of how stellar perturbations act on the Oort cloud, and to infer some of the cloud properties. Using such an approach it is shown here that the cloud radius is $\sim 10^5$ AU, the mean 'thermal' velocity in the cloud is $\sim 110 \text{ m s}^{-1}$, and the resulting perihelion distribution in the planetary region is uniform with q . Stars passing through the Oort cloud over the history of the Solar System have ejected at least 9% of the initial population, and have randomized the orbits of the remaining comets so as to leave little record of their initial state.

Dynamic studies using analytical², statistical³, and Monte Carlo⁴ techniques have shown that the cometary cloud hypothesis can explain the observed distribution of cometary orbits. While these studies have tended to emphasize the orbital evolution of the long-period comets once they enter the planetary region from the cloud, relatively little attention has been given to the dynamics of comets in the cloud, controlled principally by perturbations from passing stars. Previous studies of the stellar perturbation problem^{1,5-9} have demonstrated that no simple, complete solution exists.

The general method of treating the problem is to assume the comet is stationary in space relative to the Sun and that stars move by at high velocity along straight line paths. This is a good approximation because comets in the Oort cloud are moving at speeds of the order of $1\text{--}100 \text{ m s}^{-1}$ while the entire Solar System is moving through the local group of stars at a velocity of 20 km s^{-1} . The Sun's gravity perturbs the passing star but the deflection from a straight line is negligible: 0.12 degrees for a star passing 0.01 pc (2,063 AU) away at a velocity of 20 km s^{-1} . The velocity change imparted to the comet by the passing star is

$$\Delta V = 2GM_s/RV_s \quad (1)$$

where G is the gravitational constant, M_s the mass of the star, R the impact parameter, and V_s the velocity of the passing star. The velocity impulse is directed towards the point at which the star makes its closest approach to the comet. For a typical $1 M_\odot$ star passing at a distance of 1 pc and a velocity of 20 km s^{-1} the velocity impulse $\Delta V = 43 \text{ cm s}^{-1}$.

The Sun receives a similar impulse from the passing star though typically different in magnitude and direction. It is the net difference between the velocity impulse on the comet and the impulse on the Sun that determines the perturbation of the comet's orbit relative to the Sun. Previous studies have pointed out the importance of close stellar encounters where the comet-star distance is comparable with or less than the comet-Sun distance. The velocity changes from distant encounters are small in magnitude and tend to impart nearly the same impulse to the Sun and the comet, yielding little net effect. Close encounters, however, show much greater variation in both magnitude and direction of the velocity impulse on the comet relative to the Sun, and thus are dominant in determining the total perturbation on the comets, even though they are less frequent than the distant encounters.

The frequency of encounters between the Solar System and passing stars can be found by considering the volume of space swept out by the Solar System in a time T

$$N = \pi R^2 VT \rho_s \quad (2)$$

where N is the number of encounters, V is the velocity of the Sun relative to the local group of stars, ρ_s is the local density of stars, and R is the radius of the Solar System cross-section to be considered. Taking $V = 20 \text{ km s}^{-1}$ and $\rho_s = 0.08 \text{ pc}^{-3}$ (from ref. 10) then

$$N = 5.1 \times 10^{-6} R^2 T \quad (3)$$

where R is in parsecs and T in years. For $R = 1 \text{ pc}$ there is on the average one stellar encounter every $2.0 \times 10^5 \text{ yr}$ or 5.1 encounters per Myr. If these parameters have been constant over the history of the Solar System then the cometary cloud has experienced 2.3×10^4 stellar encounters at a distance of 1 pc or less in that time.

The mean energy perturbation caused by the passing stars can be found by integrating over concentric shells around the comet in space, with the number of stars passing at any distance R proportional to the volume of the shell at that distance.

$$\overline{\Delta V^2} = \int_{R_1}^{R_2} (\Delta V)^2 4\pi R^2 dR / \int_{R_1}^{R_2} 4\pi R^2 dR \quad (4)$$

Substituting for ΔV in equation (1) and performing the integration gives

$$\overline{\Delta V^2} = 12 (GM/V)^2 (R_2 - R_1) / (R_2^3 - R_1^3) \quad (5)$$

The lower limit of the integration can be taken to be that distance at which a single encounter will eject a comet from the cloud. At $5 \times 10^4 \text{ AU}$ from the Sun the escape velocity, $V_e = 188 \text{ m s}^{-1}$ and the impact parameter $R_1 = 472 \text{ AU}$ for a $1 M_\odot$ star passing at 20 km s^{-1} ; for 10^5 AU from the Sun $V_e = 133 \text{ m s}^{-1}$ and $R_1 = 667 \text{ AU}$. For the upper limit it has already been stated that R_2 must be of the order of the dimensions of the cometary cloud or larger: $\geq 10^5 \text{ AU}$. Thus $R_2 \gg R_1$ and the expression above can be simplified to

$$\overline{\Delta V^2} \approx 12 (GM/VR)^2 \quad (6)$$

where R is now equal to R_2 , the upper limit of the Solar System radius of cross-section being considered. Again taking $V = 20 \text{ km s}^{-1}$ and assuming a $1 M_\odot$ stars gives

$$\overline{\Delta V^2} = 0.55/R^2 \quad (\text{m s}^{-1})^2 \quad (7)$$

where R is in parsecs.

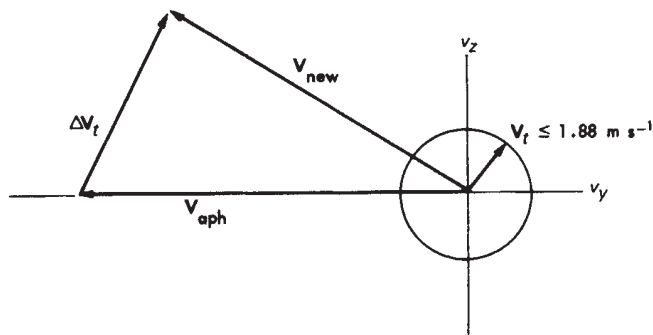


Fig. 1 Velocity phase space normal to a comet's radius vector showing relationship between initial aphelion velocity V_{aph} , tangential perturbation ΔV_t , and new velocity V_{new} . Comets falling within the 1.88 m s^{-1} circle around the origin will have perihelia $\leq 5 \text{ AU}$. An aphelion distance of $5 \times 10^4 \text{ AU}$ is assumed.

The total r.m.s. velocity perturbation in any period of time is then given by the square root of the number of stellar passages within distance R times the mean energy perturbation

$$\Delta V_{r.m.s.} = (N \cdot \overline{\Delta V^2})^{1/2} = 1.7 \times 10^{-3} T^{1/2} \text{ m s}^{-1} \quad (8)$$

where T is the time in yr. This could result in a significant perturbation on comets far from the Sun. The aphelion velocity of an Oort cloud comet with perihelion distance of 1 AU is $\sim 0.84 \text{ m s}^{-1}$ while its period is 4×10^6 yr, most of which is spent near aphelion. During that time the mean stellar perturbation from the above equation would be 3.4 m s^{-1} , capable of raising the comet's perihelion distance to over 25 AU (assuming the velocity perturbation is parallel to the aphelion velocity), outside the observable region. The angular orbit elements would also be expected to change radically. But the maximum possible change in semi-major axis for the same typical perturbation is only 0.05%. Thus comets which leave the planetary region on orbits with large aphelion distances may again be subject to significant changes in angular momentum by passing stars, but not in orbital energy. This could be an important factor in following the evolution of such comets during their lifetimes. The fraction of comets for which this occurs, however, is not large. For comets originating in the cloud and with perihelia inside the orbit of Jupiter, only $\sim 9\%$ will be returned to orbits with aphelion distances between 10^4 and 10^5 AU.

It is worthwhile to consider the total velocity perturbation on comets in the cloud over the history of the Solar System. From equation (8) $\Delta V_{r.m.s.} = 113 \text{ m s}^{-1}$, which agrees well with previous results^{1,7}. This is equal to the circular orbital velocity at 6.9×10^4 AU from the Sun or the escape velocity at twice that distance. Thus one would expect the present dynamical limits on the cloud to be of the order of 10^5 AU or less. Note that the dimensions of the cloud will shrink with time as the stellar perturbations continue to strip away the outermost shells. The mean aphelion distance of observed new comets entering the planetary region from the cloud has been shown¹¹ to be $4.32 \pm 0.12 \times 10^4$ AU. The dynamical limit¹² on the Sun's sphere of influence due to perturbations by the galactic nucleus is of the order of 2×10^5 AU.

The perihelion distribution of comets entering the planetary region from the Oort cloud can be obtained by considering the velocity phase space in the plane perpendicular to each comet's heliocentric radius vector, shown in Fig. 1. The component of the comet's aphelion velocity in this plane determines the perihelion distance. Comets are randomly scattered on this plane by stellar perturbations whose r.m.s. magnitude is $(2/3)^{1/2}$ that given by equation (8). If a comet's initial tangential velocity is given by V_{aph} then after a stellar perturbation of ΔV_t it will have some velocity V_{new} . Those comets whose new velocity vectors fall within a relatively small circle around zero transverse velocity (1.88 m s^{-1} for a perihelion of 5 AU and aphelion of 5×10^4 AU) will enter the planetary region and possibly be observed. The typical value for ΔV_t is of the order of 2.8 m s^{-1} so comets can be expected to be randomly scattered over the near-zero velocity region in phase space. Thus the number in the circle is proportional to the area of the circle, or to its radius squared, that is $\propto V_t^2$. But $V_t \propto q^{1/2}$ so the cumulative number of comets between zero and any value q is proportional to q , or dN/dq is constant. This result has also been demonstrated by Monte Carlo simulations of the problem¹³.

Another implication drawn from the magnitude of the total r.m.s. perturbation is that the orbits of the comets in the cloud have been 'randomized' over the history of the Solar System. Thus it is unlikely that orbital data could be used to discriminate between various theories of cometary origin, that is, formation among the outer planets with subsequent ejection to the cloud, versus formation in satellite fragments of the primordial solar nebula.

Finally, one can also consider the number of comets ejected from the cloud due to single passes of stars through it. Using the estimated population of 2×10^{11} comets and assuming a radius of 10^5 AU yields a number density of $\rho_c = 4.8 \times 10^{-5} \text{ AU}^{-3}$. Each

star will eject from the cloud comets for which the velocity perturbation is greater than the escape velocity:

$$2 GM_s/DV_s > (2GM_\odot/R)^{1/2} \quad (9)$$

where D is the minimum comet-star distance and R is the comet-Sun distance. Solving for D gives

$$D = (2GR/M_\odot)^{1/2} M_s/V_s \quad (10)$$

Assuming a $1 M_\odot$ star and a relative velocity of 20 km s^{-1} one finds $D = 2.1 R^{1/2}$ where R is in AU.

From equation (3) there would be $N_e = 5.4 \times 10^3$ stellar encounters within 10^5 AU over the history of the Solar System. Approximating the volume swept out by each star by a cylinder with radius equal to D at $R = 5 \times 10^4$ AU and with length equal to the mean chord length through the cloud of 10^5 AU gives

$$N_{\text{ejected}} = N_e \pi D^2 10^5 \rho_c = 1.8 \times 10^{10} \quad (11)$$

This is 9% of the estimated cloud population, a small but significant fraction. Note that this is the fraction ejected by single stellar encounters only, and that the cumulative effect of many encounters will eject additional comets.

I thank colleagues at JPL for review of an earlier draft of this paper. This paper presents one phase of work performed at the Jet Propulsion Laboratory/California Institute of Technology under NASA contract NAS 7-100.

Received 30 June; accepted 26 September 1980.

1. Oort, J. H. *Bull. astr. Instn. Neth.* **11**, 91 (1950).
2. Whipple, F. L. *Astr. J.* **67**, 1 (1962).
3. Kendall, D. G. *4th Berkeley Symp. on Mathematical Statistics and Probability* **3**, 99 (1961).
4. Weissman, P. R. *IAU Symp.* No. 81, 277-282 (1979).
5. Öpik, E. J., *Proc. Am. Acad. Arts Sci.* **67**, 169 (1932).
6. Sekanina, Z. *Bull. Astr. Instn. Czech.* **19**, 223 (1968) and **19**, 291 (1968).
7. Faintich, M. B. thesis, Univ. Illinois (1971).
8. Rickman, H. *Bull. astr. Instn. Czech.* **27**, 92 (1976).
9. Yabushita, S. *Astr. Astrophys.* **16**, 395 (1972).
10. Allen, C. W. in *Astrophysical Quantities* (Athlone, London, 1963).
11. Marsden, B. G., Sekanina, Z. & Everhart, E. *Astr. J.* **83**, 64 (1978).
12. Chebotarev, G. A. *Soviet Astr. J.* **8**, 787 (1965); **10**, 341 (1966).
13. Weissman, P. R., *Comets, Asteroids, Meteorites: Interrelations, Evolution, and Origins* (ed. Delsemme, A. H.) 87-91 (University of Toledo Press, 1977).

Far UV photolysis of CH_4 - NH_3 mixtures and planetary studies

Alain Bossard & Gérard Toupance

Laboratoire de Physico-Chimie de l'Environnement,
Université Paris Val de Marne, Avenue du Général de Gaulle,
94010 Créteil Cedex, France

The far UV photochemistry of CH_4 and of NH_3 has been widely studied^{1,2} and the primary processes are relatively well known. In contrast no experimental study of far UV photochemistry of CH_4 - NH_3 mixtures had been published before our preliminary report³. We report here new photochemical data on the effect of the mole fraction of NH_3 on the production of N-containing organics. Photolysis of CH_4 - NH_3 mixtures at 147 nm leads to the formation of nitriles when the mole fraction of NH_3 is low and to the formation of amines when the mole fraction of NH_3 is high. Study of mixtures in which the NH_3 mole fraction is very low has been particularly emphasized. The main conclusions are that (1) important quantities of nitriles may have been photo-produced on the primitive Earth if a low partial pressure of NH_3 remained and (2) nitriles are the main N-containing organics that may be photoproduced in the atmosphere of the giant planets.

The far UV photolysis of various CH_4 - NH_3 mixtures was carried out under a total pressure of 13 mbar into a cylindrical all-pyrex reactor ($l = 35$ cm; i.d. = 4 cm) with a low-pressure Xe-lamp attached (window, MgF_2 ; $\lambda = 147$ nm, monochromatic; flux emitted = 8×10^{15} photon s^{-1}). The apparatus has been described elsewhere⁴. The reactor was evacuated to a

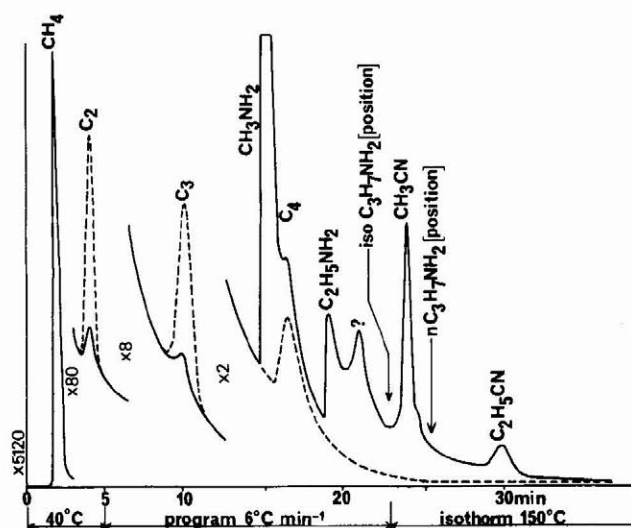


Fig. 1 Chromatogram obtained after irradiation of an equimolar $\text{CH}_4\text{-NH}_3$ mixture during 30 min with a xenon lamp emitting 8×10^{15} photons per second. Solid line, CH_4 (6.5 mbar)- NH_3 (6.5 mbar); dashed line, CH_4 (13 mbar). Chromatograms were obtained with a Varian apparatus model 1520 using FID. Chromatographic conditions, glass column (6 mm i.d. $\times$ 3 m) packed with Chromosorb 103 (100–120 mesh). Carrier gas, N_2 (40 ml min^{-1}).

pressure $< 10^{-5}$ mbar before each experiment; care was taken to ensure that there were no mercury-containing devices in any part of the apparatus. Samples were analysed mostly by gas chromatography (FID detector).

Two series of experiments were performed. In the first the irradiation time was 30 min and the mole fraction of NH_3 in the mixture ranged from 0.1 to 1; in the second the respective values were 7 min, and 0.01 to 0.1. The irradiation times were chosen to ensure a significant synthesis of products and a negligible decomposition of the initial compounds, NH_3 being the most critical, during the experiment.

The solid curve (Fig. 1) applies to an equimolar $\text{CH}_4\text{-NH}_3$ mixture irradiated for 30 min: amines, nitriles and hydrocarbons are synthesized. Gas chromatography on a Porapak Q package shows that the hydrocarbons are mainly saturated, C_2H_6 and C_3H_8 representing respectively more than 99% of the C_2 and C_3 hydrocarbons. The dashed curve applies to CH_4 irradiated for 30 min at the same total pressure. Figure 1 shows that NH_3 in the mixture strongly inhibits the synthesis of hydrocarbons. A gas chromatography on a Porapak Q package has shown that this inhibitory effect of NH_3 is much more important for unsaturated hydrocarbons than for saturated ones.

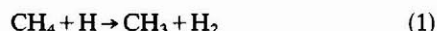
CH_3NH_2 , $\text{C}_2\text{H}_5\text{NH}_2$, HCN , CH_3CN , $\text{C}_2\text{H}_5\text{CN}$ and many aliphatic hydrocarbons have been definitively identified as products of the photolysis. As quantitative gas chromatographic analysis of HCN is not possible in the presence of NH_3 , a colorimetric titration⁵ has been achieved for this compound. For a $\text{CH}_4\text{-NH}_3$ mixture where the mole fraction of NH_3 is 0.1, the amount of HCN synthesized has been found to be approximately equal to the amount of CH_3CN synthesized during the same experiment. There has not been a search for hydrazine.

Figure 2 shows how the mole fraction of NH_3 affects the synthesis of CH_3CN and CH_3NH_2 . An opposite effect is evident: increasing the mole fraction of NH_3 strongly increases the synthesis of CH_3NH_2 but strongly decreases the synthesis of CH_3CN . In fact, for mole fractions of $\text{NH}_3 \leq 0.2$, no CH_3NH_2 has been detected by gas chromatography. To have a better limit of detectability, a spectrophotofluorimetric analysis of CH_3NH_2 using the fluorescamine⁶ technique has been performed. For a mixture containing a mole fraction of NH_3 equal to 0.1, we have

found that the quantity of CH_3NH_2 synthesized was at best 20 nmol, that is 7% of that of CH_3CN .

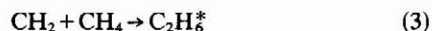
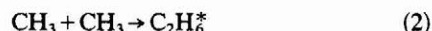
The second series of experiments, using mole fractions of NH_3 in the range 0.01–0.1, has been performed to determine more precisely the portion of the curve plotted with a dashed line on Fig. 2. The results are shown in Fig. 3. Synthesis of CH_3CN is efficient even when the mole fraction of NH_3 is very low. There was no search for amines.

The photochemistry involved in these experiments, especially for the synthesis of nitriles, is not yet clear. At 147 nm both CH_3 and CH_2 may be photoproducts from CH_4 (ref. 7) (the relative amount of the two species is not known) while both NH_2 and NH may be photoproducts from NH_3 (ref. 8). However, as molar absorption coefficients⁹ at 147 nm, are respectively $0.2 \text{ cm}^{-1} \text{ atm}^{-1}$ for CH_4 and $80 \text{ cm}^{-1} \text{ atm}^{-1}$ for NH_3 , in every considered mixture CH_4 absorbs only a minor part of the incident light (that is 2% or less when the NH_3 mole fraction is 0.1 or above). Consequently, the predominant radicals in the system are: NH_2 and NH photoproducts from NH_3 , and CH_3 mainly produced through reaction of CH_4 with energetic H atoms resulting from the photodissociation of NH_3 :



CH_2 which can be produced only from the photodissociation of CH_4 is present at a lower concentration.

The key step for the synthesis of hydrocarbons is the formation of C_2H_5^* through:



It seems likely that the reaction pathway for CH_3NH_2 synthesis combines CH_3 and NH_2 . The production of nitriles is more difficult to explain. A reaction pathway in steps must be invoked but unfortunately there is very little knowledge of the reactions of the NH radical nor of the chemistry of intermediate species which should be considered, such as CH_2NH , CH_3NH or CH_3CHNH . In fact, although the production of HCN has been reported in similar experiments¹⁰, no reaction pathway has been proposed.

Nevertheless, it seems likely that the photodissociation of CH_4 is not necessary for the production of nitriles because first, HCN is synthesized in experiments performed at 185 nm where photodissociation of CH_4 cannot occur¹⁰; and second, because of the calculated quantum yield of synthesis of CH_3CN , relative to the photons absorbed by CH_4 , obtained in our experiments.

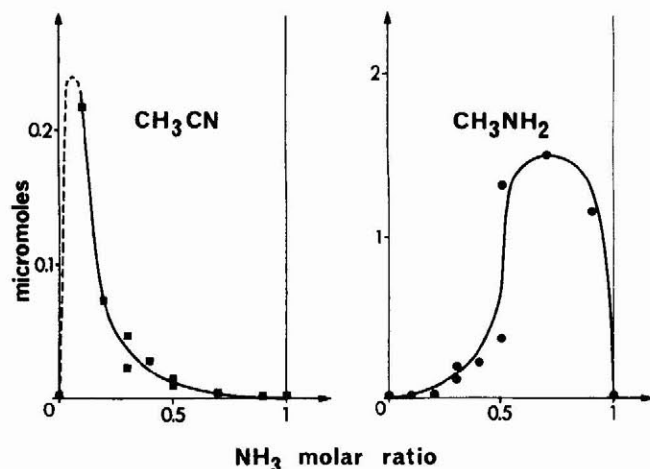


Fig. 2 Influence of the mole fraction of NH_3 on the synthesized quantities of acetonitrile and methylamine. In all these experiments, there is total absorption of the incident UV light.

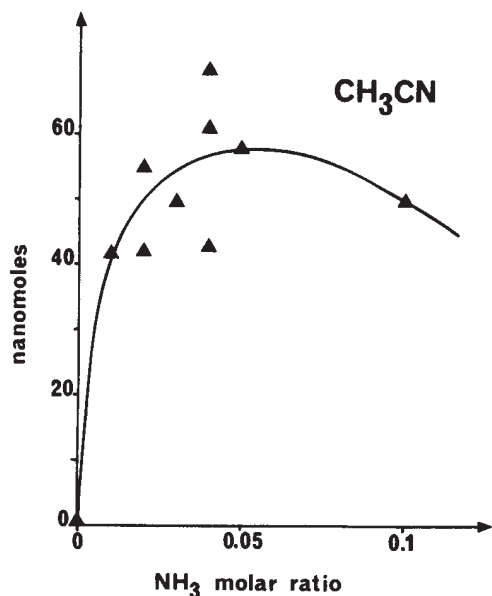


Fig. 3 Influence of the mole fraction of NH_3 on the synthesis of acetonitrile. Analysis has been performed by GC on Porapak Q support. The chromatogram obtained and analytical conditions have been previously reported⁴. Absorption of the incident UV light is not total, except when the mole fraction of NH_3 is 0.1.

Let us assume, as an extreme situation, that the photodissociation of CH_4 at 147 nm gives only CH_2 . As long as CH_4 and NH_3 remain the most concentrated species, the most frequent reactions of CH_2 are (3) and:



The kinetic parameters of reactions (3) and (4) are not known; however, as reaction (3) is very exothermic ($\Delta G^\circ = -84.4 \text{ kcal mol}^{-1}$) its probability is ~ 1 when a collision occurs. Consequently, in a CH_4 - NH_3 mixture where the mole fraction of NH_3 is 0.1, <10% of the CH_2 produced from photodissociation of CH_4 may react through reaction (4). If we assume that some of the products formed in reaction (4) are involved in the synthesis of CH_3CN , we can conclude that the quantum yield of synthesis of CH_3CN through reaction (4), calculated with respect to the photons absorbed by CH_4 , should be, at best, 0.1. It is probably very much less. As we have found that the experimental quantum yield of synthesis of CH_3CN , calculated in respect of the photons absorbed by CH_4 , is 0.4, we conclude that reaction (4) is not important for the synthesis of CH_3CN . As the second important reaction of CH_2 , reaction (3), leads to the same product as reaction (2), we also conclude that production of CH_2 is not needed to explain the bulk of the results we have obtained. Finally, as CH_3 may also be produced from reaction (1) (that is from photodissociation of NH_3) as well as from photodissociation of CH_4 , it may be assumed that photolysis of CH_4 does not introduce any significant specificity in the system.

Figure 4 represents the quantum yield of synthesis of CH_3CN and of CH_3NH_2 calculated from the data reported in Figs 2 and 3, and with respect to the photons absorbed by NH_3 during each experiment. The quantum yield of synthesis of CH_3CN is shown to increase continuously when the mole fraction of NH_3 decreases, whereas a drastically opposite effect is observed for CH_3NH_2 . Experimental constraints limited our study to mixtures in which the mole fraction of NH_3 was ≥ 0.01 . Nevertheless, we conclude that the quantum yield of synthesis of CH_3CN continues to increase slowly when the mole fraction of NH_3 is decreased to < 0.01 . A limit value of 3×10^{-2} can be accepted for very low mole fractions of NH_3 . We have also reported that

HCN is synthesized with the same efficiency as CH_3CN when the mole fraction of NH_3 is 0.1. If we assume that the same results apply at lower mole fractions of NH_3 , the limit value for the total quantum yield of synthesis of nitriles at very low mole fractions of NH_3 is probably $\sim 5 \times 10^{-2}$.

The new photochemical data on the qualitative and quantitative production of N-containing organics in CH_4 - NH_3 mixtures may be applied to the jovian troposphere and to the atmosphere of the primitive Earth.

The photochemistry of the atmosphere of Jupiter is generally approached by computer modelling based on photochemical data from CH_4 and NH_3 . Many difficulties occur in the building up of a convenient photochemical model mainly due to the lack of numerous kinetic data and even the lack of reaction pathways that would be considered. Amines are the only N-containing organics which are considered in such calculations¹¹⁻¹³. The present experimental results show that, to a large extent this assumption is wrong.

In fact at altitudes where UV light may be absorbed by NH_3 (that is middle troposphere), the NH_3 mixing ratio calculated by respect to CH_4 is $\sim 10^{-4}$. The shape of the curves in Fig. 4 suggests that the UV photosynthesis of nitriles is far more efficient than the UV photosynthesis of amines. However, this assumption must be discussed with respect to the wavelength of the UV light and to the H_2 content of the mixture.

In our experiments both NH and NH_2 are produced by photolysis of NH_3 . This is possible only for wavelengths shorter than 160 nm. At longer wavelengths only NH_2 is produced. We have previously reported⁴ the effect of important quantities of H_2 on the photochemical synthesis (147 nm) of nitriles from H_2 - CH_4 - NH_3 mixtures where the NH_3 mole fraction relative to CH_4 was 0.5. The increase of the amount of H_2 decreases the synthesis of the nitriles. We do not have similar data for amines; however, the same effect can be expected if we assume that most of the synthesis of N-containing organics is initiated by NH and NH_2 photoproducted from NH_3 . In the presence of H_2 these species are neutralized into NH_3 which should probably decrease the yield of synthesis of the different N-containing organics in similar proportions.

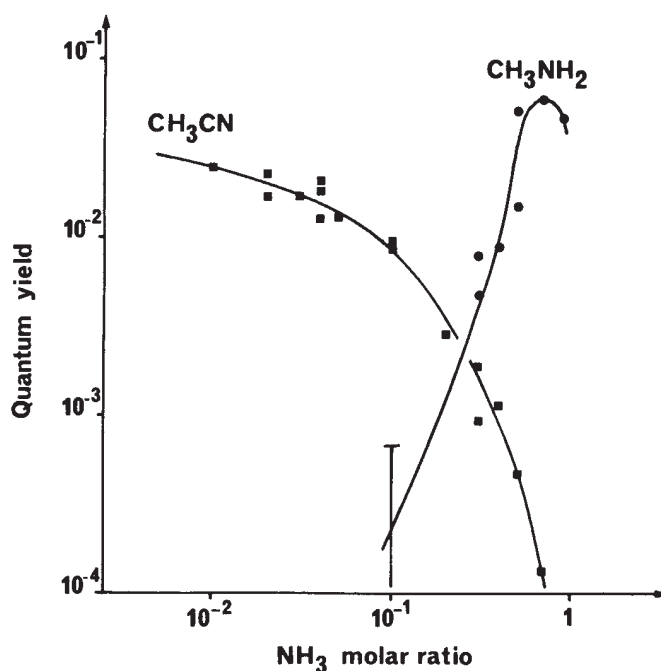


Fig. 4 Influence of the mole fraction of NH_3 on the quantum yield of formation of acetonitrile and of methylamine. These curves have been built from the experimental results shown in Figs 2 and 3. The quantum yields are calculated with respect to the photons absorbed by NH_3 .

The present data may contribute to the study of the evolution of the atmosphere of the primitive Earth in the limits of the reducing atmosphere hypothesis. They show that the presence of large amounts of NH_3 were not necessary for the photochemical synthesis of nitriles. If a low molar mixing ratio of NH_3 had remained for a significant time, an efficient production of these nitriles may have occurred. 10^{12} mol of nitriles per year would be produced only with photons of wavelengths ranging from 150 to 160 nm (quantum yield of production of nitriles = 5×10^{-2} ; present solar flux). As a cyano group cannot be easily destroyed¹⁴ by the UV light which reaches the altitude where it is produced, most of the synthesized nitriles would have reached the oceans. The presence of other gases such as N_2 or H_2O in the atmosphere should be tested. $\text{CH}_4\text{-NH}_3$ and $\text{CH}_4\text{-NH}_3\text{-H}_2\text{O}$ gaseous mixtures have been irradiated (147 nm) in the same experimental conditions: the rate of production of HCN is about the same in both cases.

This study has been supported by the French Space Agency, CNES. We thank Dr J. P. Ferris for helpful comments.

Received 1 May; accepted 3 October 1980.

- Calvert, J. G. & Pitts, J. N. Jr in *Photochemistry*, 203–204, 498 (Wiley, New York, 1966).
- Okabe, H. in *Photochemistry of Small Molecules*, 269–272, 298–299 (Wiley, New York, 1978).
- Toupance, G., Bossard, A. & Raulin, F. *Origins of Life* **8**, 259–266 (1977).
- Raulin, F., Bossard, A., Toupance, G. & Ponnampuram, C. *Icarus* **38**, 358–366 (1979).
- Bark, L. S. & Higson, H. G. *Talanta* **11**, 471–479; 621–631 (1964).
- Stein, S., Bohlen, P., Stone, J., Dairman, W. & Udenfriend, S. *Archs. Biochem. Biophys.* **155**, 202–212 (1973).
- Hellner, L., Masanet, J. & Vermeil, C. *J. chem. Phys.* **55**, 1022–1028 (1971).
- Masanet, J., Fournier, J. & Vermeil, C. *Can. J. Chem.* **51**, 2946–2951 (1973).
- Watanabe, K., Zelickoff, M. & Inn, E. C. Y. *Geophys. Res. Pap. No. 21* (1953).
- Ferris, J. P. & Chen, C. T. *Nature* **258**, 587–588 (1975); Ferris, J. P., Nakagawa, C. & Chen, C. T. *Commun. 29th COSPAR Meet. Philadelphia* (1976).
- McNesby, J. R. *J. atmos. Sci.* **26**, 594–599 (1969).
- Kuhn, W. R. & Atreya, S. K. *Geophys. Res. Lett.* **4**, 203–206 (1977).
- Atreya, S. K. & Donahue, T. M. *Rev. Geophys. Space Sci.* **17**, 388–396 (1979).
- Mizutani, H., Mikuni, H., Takahashi, M. & Noda, H. *Origins of Life* **6**, 513–526 (1975).

Trace element determination by capacitive discharge atomic absorption spectrometry

C. L. Chakrabarti, C. C. Wan, H. A. Hamed & P. C. Bertels

Department of Chemistry, Carleton University, Ottawa, Ontario, Canada K1S 5B6

Conventional instrumental analytical techniques require careful calibration of the instrument with chemically analysed standards or synthetic standards of known composition. When analyses of miscellaneous materials are required, providing the required range of standards becomes impossible. As a way to an absolute method of atomic absorption analysis, L'vov¹ suggested using a capacitive bank as a source of electrothermal energy for heating graphite atomizers. We describe here a new analytical technique which uses a capacitor bank for electrothermal heating of an anisotropic pyrolytic tube atomizer in atomic absorption spectrometry producing very fast rates of heating and an isothermal condition, both in space and in time^{1,2}, and uses a direct relationship: peak absorbance = constant $\times$ mass of analyte, thereby dispensing with analytical calibration curves. Sensitivity is almost independent of the matrix and matrix interferences are greatly reduced. Background correction is not required.

Equation (1) is a modified version of that given by L'vov³.

$$A_{\text{peak}} \approx \alpha (T^{0.7}/SP)M \approx \beta M \quad (1)$$

where A_{peak} is the peak absorbance, α a coefficient determined by atomic and spectroscopic constants and by experimental conditions, T the absolute temperature, S the cross-sectional

area, P the pressure inside the graphite tube, M is the mass of the analyte and β the proportionality constant. The apparatus and its operation have been described elsewhere¹.

Absorption pulses are recorded with a model 549 storage oscilloscope (Techtronix) fitted with a type 1A7A high-gain differential plug-in. The signal trace is photographed with a Polaroid camera (Tektronix). Three standard solutions of the analyte of different concentrations were analysed, three times each, and the arithmetic mean of these determinations was averaged to provide the value for β . The concentrations of the unknown samples were determined from their absorbance values obtained using the above conditions and suitable mass of solid samples or 5.0×10^{-6} dm³ volume of the dissolved samples as appropriate and solving equation (1) for the unknown mass of the analyte element.

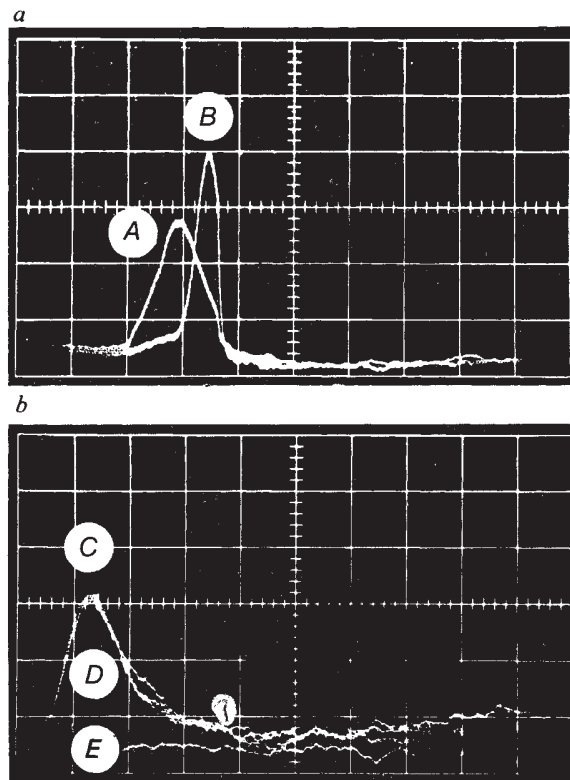


Fig. 1 *a*, Oscilloscopic traces for 5.0×10^{-13} kg of Pb (as nitrate) atomized at a heating rate of 1.3 K ms^{-1} and temperature of 2200 K. Curve A, Pb (as nitrate) in 0.5% (wt/v) NaCl aqueous solution. Curve B, Pb (as nitrate) in an aqueous solution. Vertical scale, 0.1 absorbance per scale division. Horizontal scale, 500 ms per scale division. Pb λ 283.3 nm. *b*, Oscilloscopic traces for 2.5×10^{-13} kg of Pb (as nitrate) in an aqueous solution with (curve C) and without 0.5% (wt/v) NaCl aqueous solution (curve D), atomized at a heating rate of 40 K ms^{-1} and temperature of 2,200 K. Curve E, the blank made of an aqueous solution of 0.5% (wt/v) NaCl. Vertical scale, 0.1 absorbance/scale division. Horizontal scale, 100 ms per scale division. Pb λ 283.3 nm.

Table 1 shows the experimental conditions used, and also, the precision of the proportionality constant for equation (1). From Table 1, within-the-day precision of the mean is seen to have varied from 2.5 to 7.6% r.s.d. These relatively low values of the % r.s.d. are a measure of the validity of the proportionality expressed by equation (1). The large variations in the proportionality constants between the days were mainly due to change in the resistance of the atomizer circuit caused by change in the contact resistance between graphite electrodes and graphite tubes—the change was considerable with new tubes inserted at the beginning of each day. Substitution of automatic electronic

Table 1 Experimental conditions and the precision of the proportionality constant (β) for equation (1)

Day no.	Element	Analysis line (nm)	Atomization temperature (K)	β of equation (1)	% r.s.d.*
1	Pb	283.3	2,670	5.85×10^{11}	4.3
2	Pb	283.3	2,670	7.43×10^{11}	6.6
1	Mn	279.5	2,820	7.90×10^{12}	7.6
2	Mn	403.1	2,820	8.89×10^{10}	4.0
3	Mn	403.1	2,820	1.21×10^{11}	2.5
1	Cd	228.8	2,300	2.30×10^{13}	3.5
2	Cd	228.8	2,300	9.90×10^{13}	7.6

* Three different concentrations, each concentration was determined three times, making nine determinations. The value represents % relative standard deviation of the mean of three determinations.

control of resistance for the present manual control will improve both the between-day reproducibility and the within-the-day precision—the latter is largely determined by change in resistance due to change in contact tension and in the porosity of graphite with firings.

Table 2 presents the results of analysis. Table 2*b* shows a comparison of the results obtained by the technique with those obtained by the conventional graphite furnace atomic absorption spectrometry (GFAAS) using the Perkin-Elmer Heated Graphite Atomizer (HGA) 76B and optimized experimental conditions. Table 2*b* shows that results of the conventional GFAAS suffer from extremely severe depression by the matrix, whereas the matrix interferences have been almost completely eliminated by this technique. Another problem with commercial graphite furnaces is that they are non-isothermal, both spatially and temporally⁴.

The sensitivity of the technique is almost independent of matrix. This is important because analysis by GFAAS suffers from severe matrix interferences. There is also another significant difference—background correction is not required with this technique. Because analysis of complex materials by the conventional GFAAS usually requires removal of interfering matrices (including those that give extremely intense background absorption) from samples before the determination of the analytes, background correction represents a serious limiting factor in the speed and accuracy of analysis by the conventional GFAAS. Even removal of the bulk of the interfering matrices in the charring (pyrolysis) stage of the heating cycles of the conventional GFAAS by using the time-resolved selective volatilization and atomization technique^{5,6} requires precise determination of the heating programme to remove the bulk of the matrix without any loss of the analytes, and correction for the residual background absorption by the background correction technique. The solid-sampling technique⁶, applied with this technique eliminates the large dilution factor involved in the solution-sampling technique, and thereby enhances greatly the relative sensitivity. This is in addition to the large enhancements in the absolute sensitivity of this technique, for example, for Cu, Ni and Al, the enhancements are 27-fold, 24-fold and 20-fold, respectively².

The effect of high heating rates and isothermal atomization is seen from Fig. 1*a*, which shows oscilloscopic traces of 5×10^{-13} kg of Pb atomized at a heating rate of 1.3 K ms^{-1} and an atomization temperature of 2,200 K, using the conventional GFAAS but with an anisotropic pyrolytic graphite tube. In the chloride matrix the lead signal appears earlier in time and, hence, the lead appearance temperature is lower; also the pulse amplitude is lower than that of the aqueous solution of lead nitrate without the matrix. Figure 1*b* shows oscilloscopic traces for the above systems using the capacitive discharge technique, a heating rate of 40 K ms^{-1} , and isothermal atomization at a

Table 2 Recoveries: *a*, solid sampling; *b*, solution sampling capacitive discharge technique

a, Solid sample. National Bureau of Standards oyster tissue, SRM 1566

Element	Analysis line (nm)	Certified value (kg per kg)	Recovered value (kg per kg)*	Recoveries (%)†
Pb	283.3	$(4.8 \pm 0.3) \times 10^{-7}$	$(4.7 \pm 0.10) \times 10^{-7}$	98.5
Cd	326.1	$(3.2 \pm 0.4) \times 10^{-6}$	$(3.4 \pm 0.22) \times 10^{-6}$	105
Mn	403.1	$(1.75 \pm 0.06) \times 10^{-5}$	$(1.72 \pm 0.06) \times 10^{-5}$	98.3

* The $\pm$ values represent one standard deviation of five successive replicate determinations.

† Arithmetic means of five successive replicate determinations.

b, Synthetic aqueous samples

Element	Analysis line (nm)	Mass of analyte	Matrix in excess over the analyte		Recoveries (% of the amount of analyte taken)	
					Conventional technique using HGA 76B*	New technique†
Cd	228.8	1.0 × 10 ⁻¹⁴ kg	NaCl	700,000 fold	37	100
			+MgCl ₂	84,000 fold		
			+CaCl ₂	28,000 fold		
Pb	217.0	5.0 × 10 ⁻¹³ kg	NaCl	30,000 fold	12	100
			+MgCl ₂	3,600 fold		
			+CaCl ₂	1,200 fold		
Mn	279.5	2.0 × 10 ⁻¹³ kg	NaCl	25,000 fold	75	100
			+MgCl ₂	3,000 fold		
			+CaCl ₂	1,000 fold		

* HGA 76B is Heated Graphite Atomizer 76B used with an atomic absorption spectrophotometer, model 603 (Perkin-Elmer). The relative standard deviation of the mean of five replicate determinations by the conventional graphite furnace atomic absorption spectrometry using HGA 76B = 8%.

† The relative standard deviation of the mean of five replicate determinations by the capacitive discharge technique = 12%.

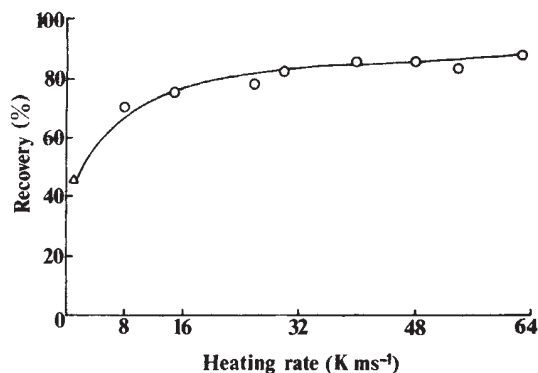


Fig. 2 Recoveries of Cd as a function of heating rate from an aqueous solution containing NaCl 1.0×10^{-8} kg + MgCl₂ 1.2×10^{-9} kg + CaCl₂ 4.0×10^{-10} kg as a function of the heating rate. Cd λ 228.8 nm. Cd = 4.0×10^{-14} kg. Atomization temperature, 1,900 K.

temperature of 2,200 K (same as that of Fig. 1a); curve E is for the blank containing an aqueous solution of 0.5% (wt/v) NaCl, and shows that the background absorption due to the NaCl matrix is completely absent, which is highly significant. Figure 1b shows that the two absorption pulses are identical in both amplitudes and areas, and the matrix effects have been completely eliminated. In Fig. 1b the lead signal appears much later in time (compared with Fig. 1a), and hence, the appearance temperature for lead is considerably higher than that in Fig. 1a. In Fig. 1b the considerably higher appearance temperature for Pb and atomization in constant temperature are crucial for eliminating chloride matrix interferences. In Fig. 1a, atomization of lead occurred in non-isothermal conditions at a temperature which was considerably lower than that in Fig. 1b and at a time when the temperature of the atomizer increased at a very rapid rate from 600 to 2,200 K; hence, the pressure inside the graphite tube changed from 1 to 3.7 atm, which resulted in partial expulsion of the lead chlorides vapour; also the extent of dissociation of lead chlorides to lead atoms was less at the lower temperature. It is concluded that the elimination of the matrix effect in Fig. 1b is due to atomization at a constant temperature and at a much faster rate of heating with the technique.

Figure 2 shows the recoveries of Cd from an aqueous solution of NaCl + MgCl₂ + CaCl₂ matrices as a function of heating rates at the isothermal atomization temperature of 1,900 K. Even at the highest heating rate used (63 K ms⁻¹), the recovery curve is seen to rise slightly, indicating a need for still higher heating rates for greater recoveries.

We conclude that even with the present laboratory-made instrumentation, the technique gives fairly accurate and precise results, which should improve with better instrumentation.

The technique with its simplified and shortened analytical procedure should save considerable time and cost per analysis. The limitation of this technique is that in the solid-sampling technique, the small sample (mg) requirement may present a sampling problem if the solid sample is inhomogeneous.

H.A.H. thanks the Government of Iraq for a postgraduate scholarship. This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada.

Received 14 April; accepted 30 September 1980.

1. L'vov, B. V. *Spectrochim. Acta* **33B**, 153–193 (1978).
2. Chakrabarti, C. L. *et al. Anal. Chem.* **52**, 167–176 (1980).
3. L'vov, B. V. *Atomic Absorption Spectrochemical Analysis*, 221 (Hilger, London, 1970).
4. Sturgeon, R. E. & Chakrabarti, C. L. *Prog. Anal. Chem. Spectrosc.* **1**, 1–199 (1978).
5. Nakahara, T. & Chakrabarti, C. L. *Anal. chim. Acta* **104**, 99–111 (1979).
6. Chakrabarti, C. L., Wan, C. C. & Li, W. C. *Spectrochim. Acta* **35B**, 93–105 (1980).

Forces between mica surfaces bearing layers of adsorbed polystyrene in cyclohexane

Jacob Klein

Polymer Department, Weizmann Institute of Science, Rehovot, Israel

Direct measurements of the forces acting between polymer surface phases adsorbed at interfaces have only recently been reported^{1–3}. These were carried out in good solvents and showed that repulsive forces were acting between the adsorbed layers; these forces increased monotonically as the surfaces approached each other. In certain conditions, however, for example, those leading to the flocculation of sterically stabilized colloidal systems, one expects the surfaces bearing the adsorbed phases to attract before repelling each other^{4,5}. I have measured the forces acting between two curved mica surfaces immersed in cyclohexane at 24 °C (a worse than θ solvent at this temperature), each bearing a surface layer of adsorbed polystyrene ($M_w = 6 \times 10^5$). No forces are observed at surface separations larger than about three radii of gyration of the polymer; on closer approach a strong attraction develops between the surfaces, changing to an ultimate repulsion as the surfaces approach closer than about one radius of gyration. Between times of a few minutes and several hours the forces are stable, well behaved and reproducible.

The experimental technique (Fig. 1) is an extension of that developed by Tabor *et al.*^{6,7}, and recently modified by Israelachvili and Adams⁸, to measure (1) the forces $F(D)$ acting between two curved mica surfaces a distance D apart; (2) the mean refractive index $n(D)$ of the medium separating the surfaces.

$F(D)$ was first determined between the mica surfaces in pure cyclohexane; polystyrene solution was then added to the cell, to a final concentration of $7 \pm 2 \times 10^{-6}$ g ml⁻¹. After 10 h incubation in the polymer solution, $F(D)$ was measured both in compression and decompression; measurements were repeated in the solution over a period of some hours, following which the polymer solution was entirely removed (save for ~1 ml between the mica surfaces) and replaced by pure cyclohexane. The surface-surface forces were then measured at intervals of between 5 min and several hours, and at compression/decompression rates of between 20–60 min per cycle, over a period of 48 h. The temperature of all experiments was 24 ± 0.5 °C.

The results are shown in Fig. 2. No forces are observed as the surfaces approach from $D \approx 300$ nm down to $D \approx 60$ nm; on closer approach an attractive force arises between the surfaces, and on further decreasing D they jump from A (Fig. 2b) to a new stable position at B (such jumps are expected⁶ whenever $\partial F/\partial D > K$, the force constant of the spring supporting the lower mica surface—see Fig. 1). Further approach results in a strong repulsion. On separation (decompression) the surfaces jump from C to a new stable position at E. The region AC is a region of instability ($\partial F/\partial D > K$) where $F(D)$ and $n(D)$ cannot be determined⁸. The 'jumps' (A $\rightarrow$ B, C $\rightarrow$ E) are not instantaneous, and up to 300 s are required for new stable positions to be established: this is interpreted as the time taken by the opposing adsorbed layers to interdiffuse (on compression) or disentangle (on decompression). The forces are stable, and within the estimated errors, reversible and reproducible over the several compression/decompression cycles. In particular, the surfaces could be held in a strongly repulsive region around $D = 14$ nm under a fixed force for up to 3 h with no resulting change in the separation.

The refractive index $n(D)$ of the medium separating the mica surfaces varied between $n = 1.56 \pm 0.02$ at $D = 14$ nm and $n = 1.43 \pm 0.012$ for $D \geq 80$ nm (Fig. 2c). The values of n were determined, following incubation, in the polymer solution, and were unchanged on replacing the solution by pure solvent, and up to 48 h later. From the bulk refractive indices of cyclohexane (1.426) and polystyrene (1.595) the value of n ($D = 14$ nm) corresponds to an adsorbance on each mica surface of around 6 mg m^{-2} of polymer. This compares with equilibrium adsorbances of $3\text{--}8 \text{ mg m}^{-2}$ for polystyrene adsorbed onto various substrates from cyclohexane at the θ temperature⁹.

In the conditions of only moderately high polymer molecular weight and extremely low bulk concentration of polymer used in the present study, the polymer solution, though considerably below the Flory θ -temperature for this system (35°C), is above its cloud-point temperature^{5,10}. Thus no phase-separation occurs¹⁰ (other than adsorption); polymer dimensions are expected to be somewhat smaller than their unperturbed value^{5,11}. Because very long times may be required before an adsorbed layer reaches equilibrium, we cannot be sure that, given much longer times, the system may not undergo further change. The fact that replacing the polymer solution by essentially pure solvent does not significantly change the adsorbed layer suggests that the adsorption may behave irreversibly. The onset of interaction between the surfaces at $D \approx 60$ nm indicates that significant overlap between the opposing polymer layers begins at this value of D , corresponding to adsorbed layers extending some 30 nm from each mica surface. This compares with an unperturbed radius of gyration $R_g = 21$ nm for this polymer¹³, and with ellipsometrically⁹ and viscometrically¹² measured equilibrium layer thicknesses of 40 nm for poly-

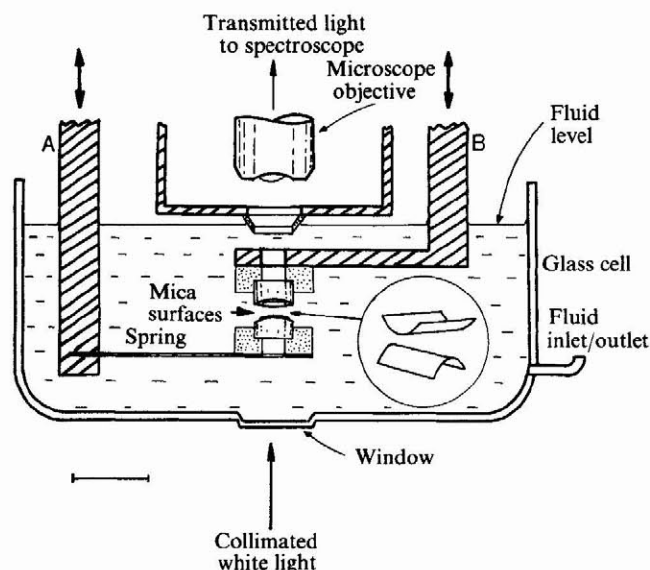


Fig. 1 Section of the apparatus used to measure forces between mica surfaces in a liquid medium. The surfaces are mounted in a crossed-cylinder configuration (inset) on cylindrical quartz lenses, within a glass cell (volume ~ 150 ml); their separation D is controlled (to 0.2 nm) by the rigid arms A, B, using a three-stage mechanism (two micrometers and a piezoelectric crystal) similar to that previously described⁸. The method is based on an optical technique utilizing white-light multiple-beam interferometry, which can measure both D (to 0.3 nm) and the refractive index $n(D)$ of the medium separating the surfaces, as well as their radii of curvature^{6,7}. The top mica surface is directly mounted on the rigid arm B, while the bottom surface is mounted on A through a leaf spring of force constant K (75 N m^{-1}). By measuring the change in D in response to known, applied relative displacements of the rigid arms A and B, $F(D)$ may be determined⁸. The glass cell is mounted within an airtight stainless steel box, which is in turn placed on a mechanically insulated platform within a thermally insulated chamber. Scale bar, 2 cm.

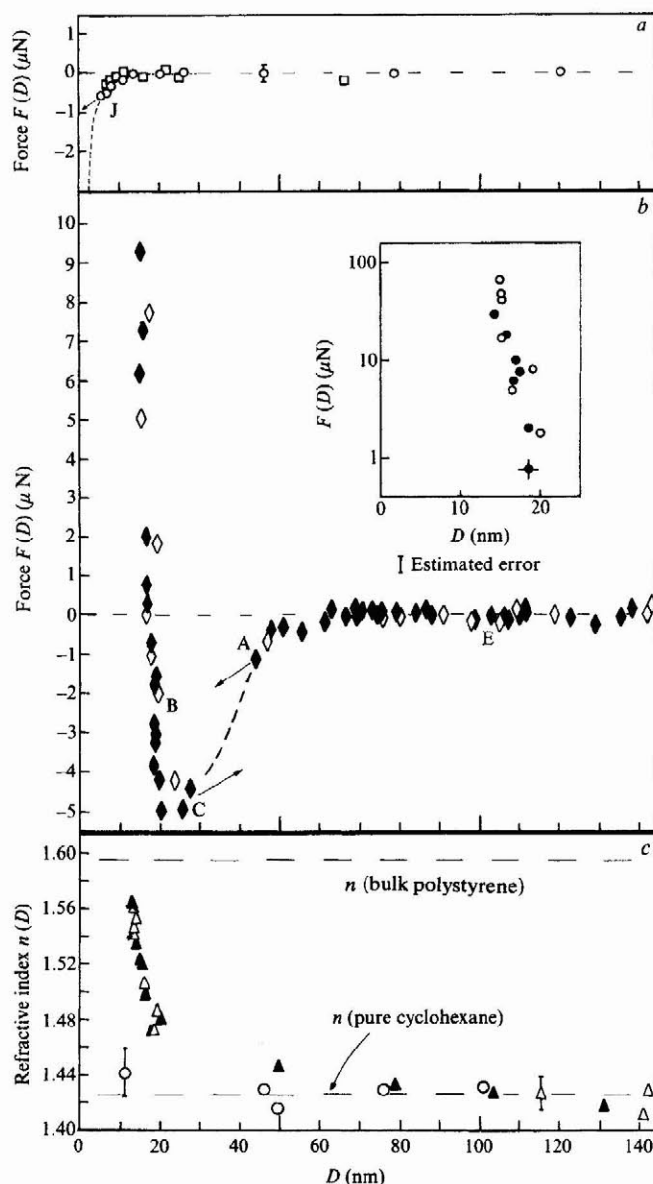


Fig. 2 Forces $F(D)$ and refractive index $n(D)$ between curved mica surfaces in cyclohexane (Fluka, spectroscopic grade) at $24.0 \pm 0.5^\circ\text{C}$. *a*, Forces between bare mica surfaces in pure cyclohexane. At $J(7 \pm 1 \text{ nm})$ the surfaces jump into contact within ± 0.3 nm of their air contact position. Mica surfaces (where R is the mean radius of curvature of mica): $\square$, $R = 0.35 \times 10^{-2} \text{ m}$; $\circ$, $R = 0.66 \times 10^{-2} \text{ m}$. The broken line is the theoretical Van der Waals force $F(D) = -AR/6D^2$, with $R = 0.66 \times 10^{-2} \text{ m}$ and a Hamaker constant $A = 1.4 \times 10^{-20} \text{ J}$, estimated from optical data for mica in cyclohexane. *b*, Forces between the curved mica surfaces of *a* ($R = 0.66 \times 10^{-2} \text{ m}$) following 10 h incubation in a polystyrene cyclohexane solution. Solutions were prepared by overnight stirring of the polystyrene (Pressure Chemicals, $M_w = 5.99 \times 10^5$, $M_w/M_n \leq 1.06$) in cyclohexane, followed by 1-h ultracentrifugation at $160,000g$ before use. The inset extends the range of $F(D)$ on a semi-log plot. $\diamond$, $\circ$, Forces measured in the polymer solution; $\blacklozenge$, $\bullet$, forces following replacement of solution by pure solvent. At A, C the surfaces jump to new stable positions B, E (see text); AC (broken curve) is a region for which $\partial F/\partial D > K$ and $F(D)$ cannot be measured⁸. *c*, Refractive index $n(D)$, of the medium separating the mica surfaces. $\circ$, in pure cyclohexane. Δ , in solution, following 10-h incubation. $\blacktriangle$, 20–44 h after replacing the polymer solution by pure solvent.

styrene (of similar molecular weight) adsorbed from cyclohexane onto various surfaces in θ conditions.

The initially attractive nature of the interactions, however, is to be expected: it arises because the polymer molecules adsorbed on each mica surface have a much greater affinity for those adsorbed on the opposing surface, than they do for the solvent molecules. As Flory has shown⁵, bringing two polymer-segment distributions into overlap in worse than θ conditions (a negative second virial coefficient) results in a negative free energy change. Thus, on initial overlap, an attractive force arises between the surfaces; as they approach to much closer than a radius of gyration, the compression of the adsorbed layers by the mica surfaces leads to extensive configurational constraints on the polymer molecules⁴, resulting in the ultimate strong repulsion observed at $D \leq 18$ nm.

Over the range of times and experimental parameters described, the present results indicate that: (1) in certain conditions the steric repulsion observed when two adsorbed polymer layers are mutually compressed is preceded by an attractive region; (2) the onset of steric interaction between the adsorbed layers may take place at surface separations comparable with the polymer dimensions and with layer thicknesses measured by other means; and (3) the polymer adsorption may be an irreversible process.

I thank Professor A. Silberberg for encouragement, interest and helpful discussions; also the design and workshop staff of the Weizmann Institute for technical assistance, and Dr J. N. Israelachvili for useful suggestions. This work was partly supported by the Basic Research Division of the Israel Academy of Sciences.

Received 17 June; accepted 12 September 1980.

1. Lyklema, H. & Van Vliet, T. *Faraday Discuss.* **65**, 25 (1978).
2. Cain, F. W., Ottewill, R. H. & Smitham, J. B. *Faraday Discuss.* **65**, 33 (1978).
3. Israelachvili, J. N., Tandon, R. K. & White, L. R. *Nature* **277**, 120 (1979); *J. Coll. Interface Sci.* (in the press).
4. Vincent, B. *Adv. Coll. Sci.* **4**, 193 (1974).
5. Flory, P. J. *Principles of Polymer Chemistry* (Cornell University Press, Ithaca, 1953).
6. Tabor, D. & Winterton, R. H. S. *Proc. R. Soc. A* **312**, 435 (1969).
7. Israelachvili, J. N. & Tabor, D. *Proc. R. Soc. A* **331**, 19 (1972).
8. Israelachvili, J. N. & Adams, G. E. *JCS Faraday I* **74**, 975 (1978); *Nature* **262**, 774 (1976).
9. Stromberg, R. R., Tutas, D. J. & Passaglia, E. *J. phys. Chem.* **69**, 3955 (1965).
10. Schultz, A. R. & Flory, P. J. *J. Am. chem. Soc.* **74**, 4760 (1952).
11. Outer, P., Carr, C. I. & Zimm, B. H. *J. chem. Phys.* **18**, 830 (1950).
12. Rowland, F., Bulas, R., Rothstein, E. & Eirich, F. *Ind. Eng. Chem.* **57**, 49 (1965).
13. Brandrup, J. & Immergut, E. H. (eds) *Polymer Handbook* 2nd edn (Wiley, New York, 1975).

Deep-seated iron ores from banded-iron formation

R. C. Morris, M. R. Thornber & W. E. Ewers

Division of Mineralogy, Institute of Earth Resources, CSIRO, Wembley, Western Australia 6014, Australia

Very large iron ore deposits have formed in many parts of the world, evidently by the supergene alteration of Precambrian banded-iron formation (BIF). Some of these deposits extend to great depths, ranging to 2,400 m at Krivoyrog¹, beyond the likely reach of oxygenated water. We propose here, on the basis of studies in the Hamersley Ranges, Western Australia, a mechanism for deep-seated ore formation in which electronic conduction through the magnetite layers in BIF connects a cathodic region near the surface, where oxygen is reduced, to an anode at depth, where iron (II) from magnetite, carbonates and silicates is oxidized and precipitated as iron (III) hydroxy-oxides. The electrical circuit is completed by ionic conduction through groundwaters. The model is based broadly on a mechanism suggested by Sato and Mooney² to explain self-potentials associated with sulphide ore bodies. It is based more specifically on a model demonstrated by Thornber³ for the

weathering of massive nickel-iron sulphide deposits at Kambalda, Western Australia, but with substantial differences in the physical situation and the reactions at depth.

The iron ore reserves of the Hamersley Iron Province exceed 33×10^9 tonnes, much of it contained in huge ore bodies, some of them exceeding 5 km along strike and extending below the present water table to as much as 400 m down dip. They conform to the bedding of the host BIF and, in common with many ore bodies of this kind throughout the world, grade from >60% Fe in ore to <30% Fe in BIF over distances that are always negligible in relation to the size of the ore body and sometimes much less than a metre.

BIF in the Hamersleys is characterized by compositional layering of the centimetre scale in mesobands, some of them magnetite-rich, alternating with others containing chert, haematite, magnetite, silicates and carbonates in varied proportions. Many of these mesobands can be traced individually for hundreds of kilometres⁴.

Morris⁵ has provided petrographical evidence that the Hamersley orebodies, in general, have formed from BIF by a combination of residual enrichment of the original iron oxides (magnetite, now martite, and primary haematite) by removal in solution of silica and other gangue, together with replacement of part of this gangue by hydrous iron oxide.

A key factor in the formation of ore at depth by such a process is the electrical conductivity of the magnetite-rich layers in BIF, and this may explain Dorr's⁶ observation that ores "form best from the oxide facies iron-formation in most places of the world".

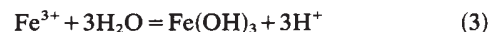
Table 1 lists some field and laboratory resistivity measurements which indicate that BIF is highly conductive along the layers of magnetite. This conductivity is retained even where the magnetite has been partially altered by weathering. Thus the BIF can be considered as an adequate conductor to allow higher oxidation potentials to penetrate to considerable depths as shown in Fig. 1a. The cathode, with a potentially vast area exposed along strike, comprises magnetite in contact with waters oxygenated by the atmosphere. Here electrons drawn from below would be consumed by the reduction of oxygen, summarized by the reaction:



The reaction that can take place at anodic surfaces of magnetite at depth, remote from sources of atmospheric oxygen, is:



The supply of Fe^{2+} to the anode, mainly from the dissolution of iron (II) minerals, all of the carbonates and silicates, and some of the magnetite, is facilitated by the release of acid in the hydrolysis reaction which will occur close to the anode surface:



The chief apparent difficulty with this model is that galvanic processes would normally favour near-surface oxidation, that is, with the anode and cathode close together; for example, where an iron bar immersed in damp soil corrodes through near the soil-air interface. However, our experimental data, as well as field and mineragraphic observations, indicate that the

Table 1 Some resistivity measurements on banded-iron formation

Measured on	Resistivity (Ωm)		Anisotropy coefficient $\lambda((\rho_t/\rho_e)^{1/2})$
	Longitudinal ρ_e	Transverse ρ_t	
Cliff face at Dales Gorge			
Mostly BIF	2.7	7,574	53
BIF + shale	4.5	7,800	42
Isolated slab 2×0.9 m thick	0.87	5.5×10^5	795
Rods 1-cm ² cross-section			
Magnetite band	0.4	—	—
Chert-magnetite	740	8.6×10^5	34.2

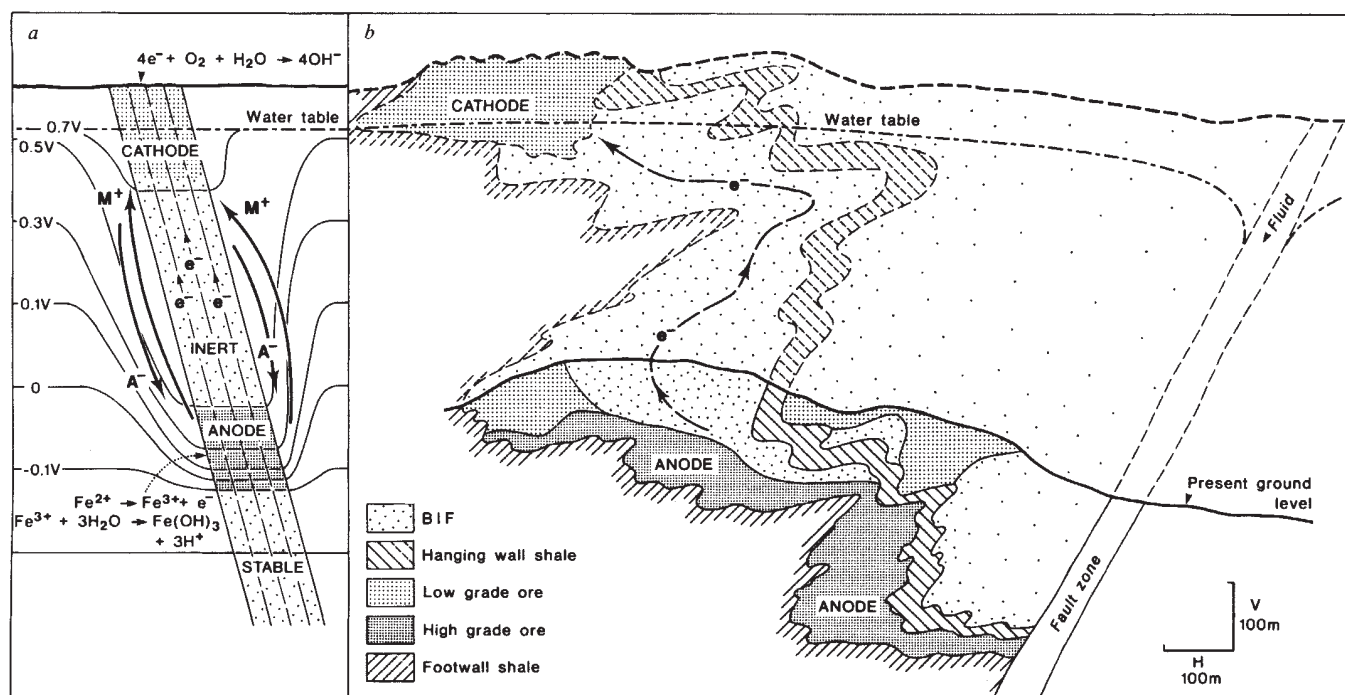


Fig. 1 *a*, Cross-section illustrating the proposed electrochemical model for the formation of iron ore at depth. Magnetite layers in the BIF provide electron paths between cathode and a deep-seated, initially porous, anodic zone. Cations (M^+) and anions (A^-) in groundwaters move through internal and external fluid pathways to complete the cell circuit. *b*, Cross-section from the Whaleback Mine at Newman, Western Australia, with a hypothetical reconstruction of the now eroded extension of the structure, illustrating the proposed electrochemical model at an advanced stage of deep iron ore formation during the Proterozoic. Fluid pathways at depth are provided initially by the fault zone, later extending laterally and vertically, by leaching of BIF.

magnetite usually weathers more slowly than other susceptible components in BIF.

Polarization tests, at pH values 4–9, on electrodes of magnetite indicate that it is an essentially inert electrode. Addition of ferrous ion to the solution produced oxidation reactions in the region of +0.3 V (SHE). With BIF electrodes the response was the same when ferrous ion was added. In the absence of added ferrous ion, however, there was still a slight indication of reaction at +0.3 V (SHE) which led us to investigate this phenomenon in a different cell.

The cell used was designed to allow current and voltage levels to be varied while surface reactions could be observed under the microscope; its design will be reported in detail elsewhere. We have shown that reactions are controlled not only by the level of conductivity of the specimens, and by the potentials to which they are subjected, but also by the proximity of reactive non-conductive minerals to the conducting bands of magnetite.

A build-up of iridescent coatings of iron oxide on magnetite surfaces is controlled by conduction, as electrically isolated grains of magnetite within the generally conductive zones are relatively unaffected, although build-up of iron oxide eventually becomes general. Silicates and carbonates in close proximity to conductive magnetites are rapidly attacked, and stilpnomelane in particular tends to swell, often to the extent that flakes of the mineral are released into the cell solution. At the same time other parts of the exposed surface become pitted as carbonate grains are dissolved; in some cases corrosion extends beyond the area of the specimen initially exposed in the cell, as solution penetrates along susceptible bands below plastic masking.

Comparable features have been observed in outcrop samples where, for example, certain bands of magnetite, carbonates and silicates may be extensively oxidized or corroded while adjacent bands are relatively fresh.

Supergene alteration of BIF does not necessarily involve oxidation of magnetite. Deep drilling in the vicinity of ore bodies at various sites in the Province has intersected severely altered and friable BIF in which chert has been strongly leached, but in

which magnetite is still fresh. In some cases this altered but unoxidized BIF grades downward into strongly goethitized zones. In the Paraburdoo 4E deposit⁷ a high-grade haematite ore lens occurs 200 m down dip below surface ore. The intervening BIF is still fresh in places despite the precambrian age⁴ of this ore.

Thus the evidence suggests that the electronic paths will be left relatively intact during deep ore formation, although leaching of other minerals takes place.

Oxidation potentials favourable for reaction (2) will exist along much of the conducting paths, but an essential requirement for deep-seated ore formation in our model is a zone at depth with greater initial porosity and better fluid access than in almost impermeable zones closer to the surface. The 'corrosion cell' is completed by electrolytic conduction through the general groundwater system, connecting the permeable anode with the cathode at the surface. These conditions would be favoured by faulting (Tom Price⁸, Whaleback⁹, Paraburdoo⁶), and by some structural situations involving damming by intrusives (Paraburdoo⁶) or cross-folding¹⁰. A hypothetical situation based on Whaleback is shown in Fig. 1*b*.

The geometry of the systems makes it difficult to test realistically the feasibility of the model in terms of rates of ore formation. A simple approach is to assume the anode and cathode to be of the same area and to regard the BIF conductor as a column of cross-section 1 m² and an appropriate length, say 400 m. Such a column will contain ~400 tonnes Fe, about one-third of it as iron (II). If a similar amount of iron (II) is introduced during ore formation, the total amount to be oxidized from iron (II) to iron (III) is about 265 tonnes. Using 4.5 Ωm as the resistivity, 0.7 V as the potential difference, and assuming the conducting path constant at 400 m, the current will be 0.39×10^{-3} A and the time to oxidize all of the iron (II) 3.7×10^7 yr. In the actual situation, we envisage that the current from a very large exposed cathode (that is the BIF outcrop), will be concentrated on a much smaller permeable zone at depth, thus increasing the current density and therefore the oxidation

rate at the anode. Even so, the calculated current density, though smaller by at least an order of magnitude than that observed in the slow corrosion of resistant alloys, such as stainless steel¹¹, is sufficient to form an ore body in reasonable geological time.

The natural BIF corrosion cell contrasts strongly, in terms of the extent of the cathode and the size of the conductor, with the situation described for Kambalda nickel ores³, in which relatively small bodies of conducting sulphides are surrounded by non-conducting country rock. A second major difference is that, whereas the conducting nickel-iron sulphides are oxidized at the anode, in the BIF it is primarily ferrous ion from the interlayered non-conducting carbonates and silicates that is oxidized as a result of the anodic reaction. This was demonstrated in the corrosion cell experiments described above.

Although the iron (II) content of the magnetite is included above in the calculation of the required current density the oxidation of magnetite, *per se*, is not considered to be essential to the first stage of ore-forming process, and will indeed inhibit it unless the oxidation forms conducting maghaemite, for which there is some evidence⁵. Ultimately much of the magnetite will be converted to haematite by a slow solid-state process, or to goethite by hydration⁵. Initial permeability to groundwater is essential, and this will be extended to an enlarged porous zone as attack on carbonates and silicates proceeds. Reactions (1), (2) and (3) will persist as long as there is magnetite and a conducting path to the surface, and diffusion along gradients of a_{Fe}^{2+} and pH will extend the effective anodic zone. Chert is a major component of these rocks and its removal, an essential part of the ore-forming process, will be facilitated by increased permeability. This leaching is governed by silica activity in ground waters, which is independent of E_h and, within normal levels, of pH.

The evidence that some iron has been added to the ore from external sources⁵ requires that the deep groundwater which transported it was anoxic. If an electrochemical mechanism were not responsible for precipitating this iron, one would need to invoke a confluence with oxygenated water from the surface which, at the depth of 2,400 m at Krivoyrog or >400 m in the Hamersley Province, would call for remarkable hydrological regimes. James *et al.*¹¹ invoked oxygenated deep artesian water during extremely arid conditions, but they pointed out that the best ore development seemed to occur in the zones most likely to remain stagnant in their model, suggesting that some other process was operating.

Finally, we stress that the electrochemical model we have proposed is an attempt to explain the more deep-seated ore bodies, and we do not doubt that many ore bodies have resulted from the penetration by oxygen-bearing surface water and, with seasonally fluctuating water tables, from processes akin to lateritization^{6,13}. These processes are to some extent electrochemical. Many such ore bodies in the Hamersley Iron Province are closely related to the Tertiary erosion surface, and some indeed form a capping⁷⁻⁹ on the deeper ore bodies. This capping probably represents the final stage in which downward progression of surface enrichment and erosion meets the upward extension of the deep-seated processes.

Received 16 April; accepted 24 September 1980.

- Belevtsev, Y. N. in *Proc. Symp. Genesis of Precambrian Iron and Manganese Deposits*, 167-180 (Unesco, Paris, 1973).
- Sato, M. & Mooney, H. M. *Geophysics* **25**, 226-249 (1960).
- Thornber, M. R. *Chem. Geol.* **15**, 1-14 (1975).
- Trendall, A. F. & Blockley, J. G. *Bull. 119. Geol. Surv. West. Australia* (1970).
- Morris, R. C. *Econ. Geol.* **75**, 184-209 (1980).
- Dorr, J. van N. II *Econ. Geol.* **59**, 1203-1240 (1964).
- Bourn, R. & Jackson, D. G. *Proc. West Australia Conf. 1979*, 187-201 (Australasian Institute of Mining and Metallurgy, 1979).
- Gilhome, W. R. *Economic Geology of Australia and Papua New Guinea*, 892-898 (Australasian Institute of Mining and Metallurgy, 1975).
- Kneeshaw, M. *Economic Geology of Australia and Papua New Guinea*, 910-916 (Australasian Institute of Mining and Metallurgy, 1975).
- Berge, J. W., Johansson, K. & Jack, J. *Econ. Geol.* **72**, 582-607 (1977).
- Jones, D. A. *Ind. Engng Chem. Prod. Res. Devel.* **11**, 12-23 (1972).
- James, H. L., Dutton, C. E., Pettijohn, F. J. & Wier, K. L. *U.S. Geol. Survey. Prof. Pap.* 570 (1968).
- Berge, J. W. *Econ. Geol.* **66**, 947-960 (1971).

A new date for the Taupo eruption, New Zealand

C. J. N. Wilson*, N. N. Ambraseys†, J. Bradley‡
& G. P. L. Walker§

* Geology Department and †Civil Engineering Department, Imperial College, London SW7 2BP, UK

‡ East Asian History of Science Library, 16 Brooklands Avenue, Cambridge CB2 2BB, UK

§ Geology Department, University of Auckland, Private Bag, Auckland, New Zealand

The Taupo eruption was a complex volcanic event that represents the most recent activity at the Taupo Volcanic Centre¹ in the North Island of New Zealand. The average² of many radiocarbon ages dates the eruption at ~AD 130. The records of ancient China and Rome refer to events in ~AD 186 of the type which follow a major volcanic eruption, and we present here reasons for considering that these events were due to the Taupo eruption, and that the true date of the eruption is ~AD 186.

The products of the Taupo eruption are rhyolitic and are dispersed over a large part of the North Island². They consist of an initial plinian pumice-fall deposit (the Hatepe Pumice)², two phreatoplinian³ ashes (the Putty Ash and Rotongaio Ash)⁴ which are separated by an erosion break, the 'ultraplinian' Taupo pumice-fall deposit⁵ and finally the non-welded Taupo ignimbrite⁶⁻⁸.

Preliminary calculations indicate an erupted volume of >60 km³ (20-25 km³ dense-rock equivalent), of which 24 km³ and >30 km³ are represented by the Taupo ultraplinian pumice-fall deposit and Taupo ignimbrite respectively. This eruption was thus comparable in magnitude to the ~1500 BC eruption of Santorini⁹.

Recent studies on the Taupo ultraplinian pumice-fall deposit⁵ and Taupo ignimbrite⁶⁻⁸ show that they were the most violent eruptions of their type yet documented. The eruption column which produced the ultraplinian pumice-fall deposit was of great power and height. Inserting the estimated dense-rock equivalent eruption rate⁵ of 10⁶ m³ s⁻¹ in published eruption-column models^{10,11} yields estimates of the eruption column height of not less than 50 km. Estimates of the initial grain-size population for the ultraplinian eruption column indicate the presence of an unusually large proportion of fine-grained material (ref. 5; and G.P.L.W. and C.J.N.W., unpublished data). It is evident that large quantities of fine ash must have been injected into the stratosphere and incorporated into the upper-atmosphere circulation.

Numerous ¹⁴C ages were determined before 1962¹² on carbonized vegetation enclosed in the eruption products, and 22 selected values were averaged by Healy² to yield the eruption-date estimate of 1,819 ± 17 yr BP, or AD 131. Note, however, that the individual ¹⁴C ages used to compile this average have much larger standard deviations (usually at least ±50 yr) than the average date itself.

A search within the ancient Chinese and Roman histories revealed the occurrence of phenomena similar to those that followed the 1883 eruption of Krakatoa¹³.

In the Chinese literature, it is recorded that:

"During the reign of (the emperor) Ling Ti (AD 168-189) several times the Sun rose in the east red as blood and lacking light, only when it had risen to an elevation of more than two *zhàng* was there any brightness. When it set in the west, at two *zhàng* above the horizon it was similarly red. . . . Also during this period, several times when the Moon rose and set and was two to three *zhàng* above the horizon, all was red as blood."¹⁴

‡ Present address: 18 Juniper Lane, Amherst, Massachusetts 01002.

A zhàng is approximately equal to 12° (ref. 15).

In the Roman literature, two sources refer to unusual appearances of the sky in, or just before, AD 186.

"Before the war of the deserter (autumn 186), the heavens were ablaze."¹⁶

"Stars were seen all the day long and that some did stretch in length hanging as it were in the midst of the ayre which was a token of a cloud not kindled but drouen together for it seemed kindled in the night, in the day when it was far off it vanished away."¹⁷

These descriptions are reminiscent of the effects observed to follow the 1883 Krakatoa eruption and particularly the characteristics of the Krakatoa dust haze (ref. 13, pp. 223–225) in higher latitudes. The appearance and description of the haze as being too thin to produce blue or green-coloured suns, as having a rippled or striated appearance and only being visible around dawn and dusk (recall "... in the day when it was far off it vanished away") and the accompanying glow effects, seem to correspond closely with the above descriptions.

These records of phenomena resulting from a major eruption in or just before AD 186 and the absence of records of similar effects between then and the Vesuvius eruption of AD 79, suggest that the records do indeed refer to the Taupo eruption and therefore that the true date of this eruption is ~AD 186.

Two problems arise. First, what chance is there of material erupted at Taupo (38° 50' S) being able to spread across the tropics and into the Northern Hemisphere in quantities sufficient to cause noticeable effects? Although Northern Hemisphere temperature records^{18,19} show a minor change at the time of the 1886 Tarawera (New Zealand) eruption²⁰, this is not considered as significant enough to indicate that material can cross the tropics. Lamb^{21,22} has stated (for example, ref. 22, p. 416) that: "... volcanic dust originating in high latitudes hardly spreads in significant quantities beyond about latitude 30° in the hemisphere of origin; though small quantities may reach any part of the Earth, it effectively covers only latitudes 30°–90° in the hemisphere of origin...". Note that Lamb was referring to the dispersal of volcanic ash in the lower stratosphere, from the tropopause up to ~25–27 km. On the other hand, data quoted earlier indicate that the Taupo ultraplinian eruption-column penetrated the upper stratosphere and reached the mesosphere. Fine ash injected at such levels would take times of the order of months to years to fall to lower levels²¹ and in the meantime be subject to the high level (>30 km) drift which crosses the equatorial regions between the hemispheres (see ref. 21, p. 455). Thus the finest ash, injected at the greatest heights, could spread from New Zealand to the Northern Hemisphere with little difficulty.

This is in accord with the above descriptions; the absence of blue or green suns or moons suggests that the causative particles were <5 µm across^{13,22}. Also, the nature of the Chinese description implies that the glow effects appeared gradually and lasted for some time; other reports of sunset-glow effects found by one of us (J.B.) in the Chinese literature (almost certainly caused by middle or northern-latitude eruptions) are generally closely dated (often to the nearest month) as to when they first appear.

The second problem is that an AD 186 eruption date lies just outside three standard deviations of the average of the ¹⁴C dates (1,819 ± 17 yr BP). However, since the dates were published, much work has been done on the corrections required to convert ¹⁴C ages to true ages derived by dendrochronology. Michael and Ralph²³ have obtained ¹⁴C dates of AD 118 and AD 148 on samples with true dates of AD 150 and AD 197 respectively, while Jess²⁴ has obtained a ¹⁴C date of AD 119 on material of true date AD 185. (All these ¹⁴C dates were calculated using the 5,568-yr half life, for valid comparison with the New Zealand dates; reported standard deviations are ±40, 44 and 45 yr respectively.) The difference between the proposed AD 186 true and AD 131 ¹⁴C eruption dates is, therefore, very close to that between true and ¹⁴C dates obtained on samples from the same true-age period.

One slight possibility is that these atmospheric phenomena resulted from the eruption which produced the north lobe of the White River Tephra in Alaska²⁵, ¹⁴C dated at about 1,890 yr BP (average of 11 determinations²⁵). Applying the true date to ¹⁴C date corrections as above^{23,24} suggests that in this case, the ¹⁴C date of ~AD 60 is close to the true age, and that the White River eruption did not produce the effects seen in or just before AD 186 here ascribed to the Taupo eruption.

We conclude that the Taupo eruption, previously dated by ¹⁴C methods at AD 131, actually occurred in ~AD 186. The scale of the eruption, its violence, and the wide dispersal of the resulting air-fall deposits suggest that enough fine dust escaped into the upper-atmospheric circulation to reach the Northern Hemisphere and produce the optical effects observed in China and Europe.

Whilst this new date is only a proposal, we hope that we have demonstrated the potential of using historical records in dating and characterizing the distant effects of large volcanic eruptions.

Confirmation of this new date for the Taupo eruption might be sought in Antarctic or Arctic ice cores. Alternatively, a New Zealand dendrochronology time scale could be constructed and correlated with the timber enclosed abundantly in the Taupo ignimbrite.

We thank Dr J. Needham for assistance and Drs J. Westgate and L. Wilson for helpful comments. C.J.N.W. is an NERC research student and G.P.L.W. a Captain James Cook Research Fellow of the Royal Society of New Zealand. J.B. thanks the East Asian History of Science Library for facilities.

Received 9 July; accepted 8 October 1980.

1. Healy, J. *Bull. volcan.* **26**, 141–151 (1963).
2. Healy, J., Vucetich, C. G. & Pullar, W. A. *Bull. N.Z. geol. Surv.* n.s. 73 (1964).
3. Self, S. & Sparks, R. S. J. *Bull. volcan.* **41**, 1–17 (1978).
4. Walker, G. P. L. *J. volcan. geotherm. Res.* (in the press).
5. Walker, G. P. L. *J. volcan. geotherm. Res.* **8**, 69–94 (1980).
6. Walker, G. P. L., Heming, R. F. & Wilson, C. J. N. *Nature* **283**, 286–287 (1980).
7. Walker, G. P. L., Wilson, C. J. N. & Froggatt, P. C. *Geology* **8**, 245–249 (1980).
8. Walker, G. P. L., Wilson, C. J. N. & Froggatt, P. C. *J. volcan. geotherm. Res.* (in the press).
9. Watkins, N. D. *et al. Nature* **271**, 122–126 (1978).
10. Settle, M. J. *J. volcan. geotherm. Res.* **3**, 309–324 (1978).
11. Wilson, L., Sparks, R. S. J., Huang, T. C. & Watkins, N. D. *J. geophys. Res.* **83**, 1829–1836 (1978).
12. Grant-Taylor, T. L. & Rafter, T. A. *Radiocarbon* **5**, 118–162 (1963).
13. Symons, G. J. (ed.) *The Eruption of Krakatoa and Subsequent Phenomena* (Report of the Krakatoa Committee of the Royal Society, London, 1888).
14. *Hou Han Shu* Vol. 11, 3373, lines 2–3. Peking (punctuated) edn (1965).
15. Kiang, T. *Mem. R. astr. Soc.* **76**, 27–66 (1972).
16. *Scriptores Historiae Augustae: Commodus Antonius* xvi, 2 (Loeb, 1967).
17. Herodian, *Meta Markon vasileias historiai* i, 10 (ed. K. Stavenhagen, 1922).
18. Humphreys, W. J. *The Physics of the Air* (McGraw Hill, New York, 1940).
19. Rampino, M. R., Self, S. & Fairbridge, R. W. *Science* **206**, 826–829 (1979).
20. Cole, J. W. N. Z. *J. Geol. Geophys.* **13**, 879–902 (1970).
21. Lamb, H. H. *Phil. Trans. R. Soc. A266*, 425–533 (1970).
22. Lamb, H. H. *Climate Present, Past and Future* Vol. 1 (Methuen, London, 1972).
23. Michael, H. N. & Ralph, E. K. *Radiocarbon* **16**, 198–218 (1974).
24. Suess, H. E. *Radiocarbon* **20**, 1–18 (1978).
25. Lerbekmo, J. F., Westgate, J. A., Smith, D. G. W. & Denton, G. H. *Quaternary Studies* (eds Suggate, R. P. & Cresswell, M. M.) 203–209 (Royal Society of New Zealand, Wellington, 1975).

Output rate of magma from active central volcanoes

G. Wadge

Lunar and Planetary Institute, 3303 NASA Road One, Houston, Texas 77058

For part of their historic records, nine of the most active volcanoes on Earth have each erupted magma at a nearly constant rate. These output rates are very similar and range from 0.69 to 0.26 m³ s⁻¹. The volcanoes discussed here—Kilauea, Mauna Loa, Fuego, Santiaguito, Nyamuragira, Hekla, Piton de la Fournaise, Vesuvius and Etna—represent almost the whole spectrum of plate tectonic settings of volcanism. A common mechanism of buoyantly rising magma-filled cracks in the upper crust may contribute to the observed restricted range of the rates of output.

Table 1 Historic output rates of magma from selected volcanoes

Volcano	Plate tectonic setting	Product	Period	Volume ($\times 10^6 \text{ m}^3$)	No. of eruptions	Output rate ($\text{m}^3 \text{ s}^{-1}$)	Refs
Kilauea	Intraoceanic	Tholeiitic basalt	1823–1975	3,323	54	0.69	1–3
Mauna Loa	Intraoceanic	Tholeiitic basalt	1823–1975	3,158	37	0.66	1, 4
Fuego	Subduction	High Al_2O_3 basalt	1932–1979	700	28	0.5	5
Santiaguito	Subduction	Dacite	1922–1972	658	14	0.42	6
Piton de la Fournaise	Intraoceanic	Hawaiite	1931–1974	529	21	0.39	7
Nyamuragira	Intracontinental	Leucite basanite	1901–1971	838	12	0.38	8
Hekla	Spreading ridge	Basaltic andesite	1693–1970	2,731	4	0.31	9
		Dacite					
Vesuvius	Intracontinental?	Leucite basanite	1760–1944	1,660	20	0.29	10, 11
Etna	Subduction	Hawaiite	1759–1974	1,746	31	0.26	12

The volumes of lava and tephra have been corrected to magmatic volumes at liquidus temperatures. The bulk densities of the lava flows are not accurately known, but the corrections are small. A factor of 0.85 was applied to the lavas of Kilauea¹⁴ and also to Mauna Loa and Hekla. A factor of 0.94 based on my unpublished data for lavas of Etna, was applied to the lavas of Etna, Nyamuragira and Santiaguito. For the dacitic tephra of Hekla a correction factor of 0.20 was used³³, and one of 0.45 was used for the basaltic tephra of Etna and Nyamuragira. The number of eruptions is a general guide, and in the cases of Etna and Vesuvius refers only to the major lava-forming eruptions. For Hekla with its long repose periods, the time used to calculate the output rates starts after the 1693 eruption and ends after the 1970 eruption. The output rate for Piton de la Fournaise as originally published⁷ was too great by a factor of 10.

The rates at which magmas have been erupted at these nine volcanoes have been calculated from estimates of the volumes of their historic products^{1–12} (Table 1). These volcanoes were chosen because of the availability of data of similar quality. Each is a major central volcano of considerable age ($>10,000$ yr) which has been active for much of the recent historic past. Using volumetric data to calculate output rates can only be justified for those volcanoes whose cumulative volume curves are approximately linear (Fig. 1). The volcanoes of Table 1 not shown in Fig. 1 also satisfy this criterion. It has been suggested that the combined output rate for Mauna Loa and Kilauea is constant¹³. Because of the uncertainties in the data for the summit, caldera-filling lavas for both these volcanoes, complete cumulative volume curves are not shown in Fig. 1. However, judged solely on the historic flank eruptions of Mauna Loa which account for $\sim 80\%$ of the total for this volcano, the curve is linear. Hence the output rate is considered valid for Mauna Loa and probably also for Kilauea.

The nine volcanoes of Table 1 show a very restricted range of output rates which cannot be considered an artefact of a single observer's bias despite the levels of uncertainty (several tens of per cent) in the volume estimates. This small sample cannot be considered representative of volcanism in general in that these volcanoes have been unusually active and unusually well studied. Nevertheless, there is no ready explanation for such a common output rate and its significance may extend beyond the individual volcanoes reported here. Departures from linearity in the cumulative volume curves of Vesuvius and Etna are apparent in Fig. 1 and have been previously discussed in terms of the evolution of intravolcanic magma chambers and magma supply^{11,12}. An episodic character to the recurrence of activity at Fuego, represented here in its most recent history has been detected⁵. Thus a constant output rate may represent a state of equilibrium that can only be maintained for a limited period of time.

Not all magmas that rise beneath volcanoes reach the surface. Some may freeze in intrusions¹⁴, or become involved in crystal fractionation and host rock digestion¹⁵. Kilauea is the only volcano for which there is geological and geophysical evidence that magma is supplied at a constant rate which is greater than the subaerial output rate¹⁴. This supply rate of $3.6 \text{ m}^3 \text{ s}^{-1}$ is about five times the observed historic output rate, the balance being taken up mainly by submarine effusion and dyke intrusion. This has been applicable for the past 30 yr (refs 14, 16, 17). Supply rates for magma entering the other volcanoes can only be guessed at. The evidence from Kilauea suggests that constant output rates at each of the nine volcanoes represent a partitioning of ascending magma into a fraction which is erupted and a fraction which remains below the surface. Thus it is conceivable that each volcano has a constant supply rate that is larger than its

output rate but falling within a similarly restricted range of values.

Buoyant rise of magma in cracks is a mechanism which may explain a constant supply rate to these volcanoes. A magma-filled crack experiences a net force of $(\rho - \rho')gV$, where ρ is the density of the rock, ρ' is the density of the magma, g is the value of gravity and V is the volume of the crack¹⁸. The densities of the magmas considered here range from $\sim 2,600 \text{ kg m}^{-3}$ (olivine tholeiite) to $2,300 \text{ kg m}^{-3}$ (dacite) at liquidus temperatures and atmospheric pressure. Up to 5 kbar pressure the increase in densities is small ($<100 \text{ kg m}^{-3}$)^{19,20}. The density contrast for basalts will be smaller than for andesites and dacites, and for an average crustal density of $2,670 \text{ kg m}^{-3}$ a factor of five covers the range of density contrasts $(\rho - \rho')$, though this is sensitive to the local rock density. V is restricted by the lithospheric stress field; the greater stress at the base closes the crack as its tip migrates upwards. A free crack has a stable vertical length, L , given by

$$L \leq 2 \left[\frac{K_c}{\sqrt{\pi} g (\rho - \rho')} \right]^{2/3}$$

where K_c is the fracture toughness of the rock²¹. For a density contrast of 100 kg m^{-3} in rocks of fracture toughness of $100 \text{ MN m}^{-2/3}$ L is $\sim 3,000$ m. The width of the crack is variable but an average width of 10 m is geologically realistic and this would give a typical volume for a disk-shaped crack of about $70 \times 10^6 \text{ m}^3$ and a net buoyancy of 70 GN.

This crack volume is of the same order of magnitude as the individual eruption volumes for most of the volcanoes in Table 1. Such cracks would need to arrive at the volcano every 6–7 yr to account for the observed output rates, more frequently if there is subsurface loss of magma. Magma-filled cracks, therefore, are of appropriate size and the ascending magma has a buoyancy controlled by parameters which may vary within sufficiently narrow limits to produce the observed similarity of output rates of volcanoes from very different tectonic settings.

If the magma is to remain liquid during conduction of heat to the cold walls it must have a minimum average ascent velocity. The calculations of Marsh for andesite-filled cracks²² give a velocity of 10^{-3} m s^{-1} . A recent quantitative model of magma ascent at Kilauea²³ involves magma rising along cracks at velocities between 10^{-2} and 10^{-1} m s^{-1} . The petrology and geochemistry of the products of the October 1974 eruption of Fuego have been interpreted²⁴ in terms of a magma-filled crack of volume $100 \times 10^6 \text{ m}^3$ rising at an average velocity of 10^{-3} m s^{-1} for the last 8 km to the surface. These figures indicate the minimum average velocities though the real motion will be incremental²⁵. In addition, the viscosity of the liquid has an important limiting effect on crack velocity²⁶. The limiting effect will be much stronger for highly viscous siliceous magmas than for that for less viscous basaltic magmas. Although this problem

cannot be considered here it has an important bearing on volcanic magma supply.

Most of the world's mature, subaerial central volcanoes have been considerably less active in historical times than those discussed here. Many have erupted no magma for hundreds of years. This may mean that magma is not being supplied to these volcanoes because it is being intercepted by a large crustal reservoir or because magma is not entering the crust at all. In a review of ash flow eruptions of all sizes, Smith²⁷ proposes a tentative correlation between ejecta volume and the time taken to produce this material. This rate is $0.03 \text{ m}^3 \text{ s}^{-1}$ ($10^{-3} \text{ km}^3 \text{ yr}^{-1}$), which is about 1/10th of the output rates presented here. There have been several published estimates of the volumetric productivity of large sections of several volcanic arcs situated above subduction zones. These estimates are usually expressed in $\text{km}^3 \text{ Myr}^{-1} \text{ km}^{-1}$ of arc, and range from 5 to $30 \text{ km}^3 \text{ Myr}^{-1} \text{ km}^{-1}$ (refs 28–32) for various periods of the Quaternary. If an average of 50 km of arc is allocated to each volcanic centre, each volcano produces $\sim 250\text{--}1,500 \text{ km}^3 \text{ Myr}^{-1}$ or $0.008\text{--}0.05 \text{ m}^3 \text{ s}^{-1}$. This is between 2 and 10% of the historic

output rates shown by the Table 1 volcanoes near subduction zones. These two comparisons suggest that if the output rates reported here are more generally applicable to arc volcanoes when they are in an active state, then their corresponding periods of inactivity must last about 10–50 times longer.

I thank W. I. Rose Jr, S. Self and C. A. Wood for reviews of an earlier manuscript. Part of this work was done while I was a visiting scientist at the Lunar and Planetary Institute which is operated by the Universities Space Research Association under Contract NSR-09-051-001 with NASA.

Received 23 July; accepted 29 September 1980.

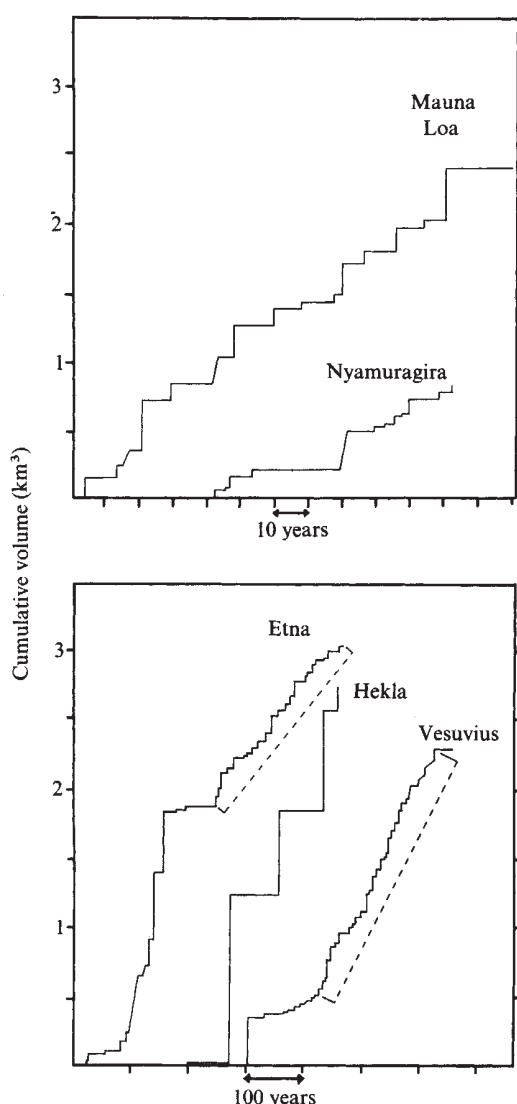


Fig. Cumulative volume curves compiled from the historic records of five volcanoes. The curve for Mauna Loa is for flank eruptions during 1843–1978. The curves for Nyamuragira and Hekla are for the periods used in Table 1, whilst the linear portions of the curves for Etna and Vesuvius used in Table 1 are bracketed by the dashed lines. The solid curves for Etna¹² are for 1535–1974, and for Vesuvius^{10,11}, 1630–1979. Note that the abscissa scale for Mauna Loa and Nyamuragira is different from that for the other three volcanoes.

1. MacDonald, G. A. *Catalogue Active Volcanoes of the World* Vol. III (1956).
2. MacDonald, G. A. & Abbott, A. T. *Volcanoes in the Sea* (University of Hawaii Press, 1970).
3. Tilling, R. I., Holcomb, R. T., Lockwood, J. P. & Peterson, D. W. *Int. Colloq. Planet. Geol.*, Rome, 149–152 (1975).
4. Lockwood, J. P., Koyanagi, R. Y., Tilling, R. I., Holcomb, R. T. & Peterson, D. W. *Geotimes* 21, 12–15 (1976).
5. Martin, D. P. *EOS* 61, 70 (1980).
6. Rose, W. I. Jr *Bull. volcan.* 37, 73–94 (1973).
7. Ludden, J. N. J. *Volcan. geotherm. Res.* 2, 385–395 (1977).
8. Poucllet, A. *Bull. volcan.* 29, 1–13 (1975).
9. Thorarinsson, S. *The Eruption of Hekla 1947–48 I* (1–170 Visindafelg Islinginga 1967).
10. Imbo, G. *Annali. Oss. vesuv.* 5, 185–379 (1951).
11. Yokoyama, I. in *Volcanoes and Tectonosphere* (eds Aoki, H. & Iizuka, S.) 93–101 (Tokai University Press, 1976).
12. Wadge, G., Walker, G. P. L. & Guest, J. E. *Nature* 255, 385–387 (1975).
13. Moore, J. G. *Bull. volcan.* 34, 562–576 (1970).
14. Swanson, D. A. *Science* 175, 169–170 (1972).
15. O'Hara, M. J. *Nature* 266, 503–507 (1977).
16. Dzurisin, D. *EOS* 61, 70 (1980).
17. Swanson, D. A., Duffield, W. A., Jackson, D. B. & Peterson, D. W. *U.S. Geol. Surv. Prof. Pap.* 1056 (1979).
18. Weertman, J. J. *geophys. Res.* 76, 1171–1183 (1971).
19. Fujii, T. & Kushiro, I. *Yb. Carnegie Instn Wash.* 76, 419–424 (1977).
20. Kushiro, I. *Yb. Carnegie Instn Wash.* 77, 675–677 (1978).
21. Muller, O. H. & Muller, M. R. *Abstr. 11th lunar planet. Sci. Conf.*, Houston, 774–776 (1980).
22. Marsh, B. D. *Phil. Trans. R. Soc. A288*, 611–625 (1978).
23. Wright, T. L., Shaw, H. R., Tilling, R. I. & Fiske, R. S. *Abstr. Hawaii Symp. Intraplate Volcanism and Submarine Volcanism*, Hilo, 108 (1979).
24. Rose, W. I. Jr, Anderson, A. T. Jr, Woodruff, G. & Bonis, S. B. J. *Volcan. geotherm. Res.* 4, 3–53 (1978).
25. Weertman, J. J. *geophys. Res.* 76, 8544–8553 (1971).
26. Anderson, O. L. *Bull. volcan.* 41, 341–353 (1978).
27. Smith, R. L. *Geol. Soc. Am. Spec. Pap.* 180, 5–27 (1979).
28. Sapper, K. *Vulkankunde* (Engelhorn, Stuttgart, 1927).
29. Nakamura, K. in *The Utilization of Volcano Energy* (ed. Colp, J. L.) 273–285 (Sandia Laboratories, Albuquerque 1976).
30. McBirney, A. R. *Can. Miner.* 14, 245–254 (1976).
31. Marsh, B. D. *Am. Sci.* 67, 161–172 (1979).
32. Sigurdsson, H., Sparks, R. S. J., Carey, S. & Huang, T. C. J. *Geol.* (in the press).
33. Thorarinsson, S. *Geol. Rdsch.* 57, 705–719 (1967).

Calcium-carbonate-phosphate surface complex in calcareous systems

Yoram Avnimelech*

Faculty of Agricultural Engineering, Technion,
Israel Institute of Technology, Haifa, Israel

The interactions between soluble orthophosphates and lime (CaCO_3) have been extensively studied because of the frequent occurrence of soil systems rich in CaCO_3 and the dominant effect of calcium carbonate on phosphate solubility. Several investigators have suggested that in these systems chemically defined calcium phosphate species are precipitated. Cole and Olsen¹ found that the solubility of phosphate in calcareous soils seems to be controlled by a di-calcium phosphate (DCP) solid phase, although the solubility product inferred (2.89×10^{-9}) is significantly less than that measured² for DCP (2×10^{-7}). Similar results have been reported for the interaction between fish-pond sediments and the aqueous phase³. On the other hand, Stum and Leckie⁴ have suggested that the solubility of phosphate in calcareous systems may be controlled by the chemisorption of phosphate on CaCO_3 particles, with the formation of amorphous calcium phosphates³ or of surface complexes.⁴ We now describe experiments which argue for the formation of a surface complex of calcium-carbonate-phosphate with well defined chemical composition.

* Present address: The USDA-SEA Water Quality Management Laboratory, Durant, Oklahoma 74701.

The experiments consisted of adding phosphate to suspensions of analytical grade CaCO_3 . In the first series, K_2HPO_4 solutions, buffered with KOH to pH 8.4, were added drop by drop. Later, phosphate-enriched anion exchange resin (Dowex 2) contained in a fine nylon cloth was used to maintain a slow supply of phosphorus. The suspensions were shaken for periods of 3–25 days. Systems planned to provide data on material balance were tightly sealed to avoid gains or losses of CO_2 . The samples were filtered and measurements made of: pH with a glass electrode, calcium with atomic adsorption using nitrous oxide flame, phosphate colorimetrically⁶ and carbonates by titration⁷. Ionic distribution in the solution was computed using constants for the dissociation of the orthophosphoric and carbonic acids and stability constants for the ionic pairs CaHPO_4 , $\text{CaH}_2\text{PO}_4^+$ (ref. 2), CaCO_3° , CaHCO_3^- (ref. 8) and CaOH^+ (ref. 9).

The material balance of components in the solution during the reaction of potassium phosphate with calcium carbonate in a sealed system over 25 days is given in Table 1. A transfer of a component from the solution to the solid is defined as a negative change whereas dissolution, or desorption, is defined as a positive change. The balance for hydrogen is made on its free ionic species, H_3O^+ , as well as on other species containing protons, such as HCO_3^- and HPO_4^{2-} .

Calcium is lost from the solution after the addition of phosphorus, but the losses become smaller with the increased phosphorus concentration and are not related to the changes in phosphorus concentration. The changes in total carbonate concentration follows the same pattern, but in the range of high phosphorus losses (above $\sim 3 \times 10^{-4} \text{ mol P l}^{-1}$), carbonate is released from the solid into the solution. The transfer of hydrogen from the solution to the solid is linearly related to that of phosphorous. The linear regression is:

$$\Delta\text{H} = 0.59 + 2.11 \Delta\text{P}, \quad r = 0.965 \quad (1)$$

and the best estimate of ΔH as a function ΔP is

$$\Delta\text{H} = 2.625 \Delta\text{P} \quad (2)$$

If the reaction occurring in the system tested here was the precipitation of any known calcium phosphate, the ratio of precipitated calcium to that of phosphorus should be constant and in the range between 1.67 to 1.0, the stoichiometric Ca/P ratios in hydroxyapatite and DCP, respectively, that is

$$\frac{\text{Ca precipitated}}{\text{P precipitated}} = \text{constant}; \quad 1.0 \leq \text{constant} \leq 1.67 \quad (3)$$

The precipitated calcium includes that disappearing from the original solution, as well as calcium that is possibly added into the solution, by a dissolution of CaCO_3 . Any gain of soluble calcium is associated with an identical gain in carbonate. Thus, calcium precipitated with phosphate is given by

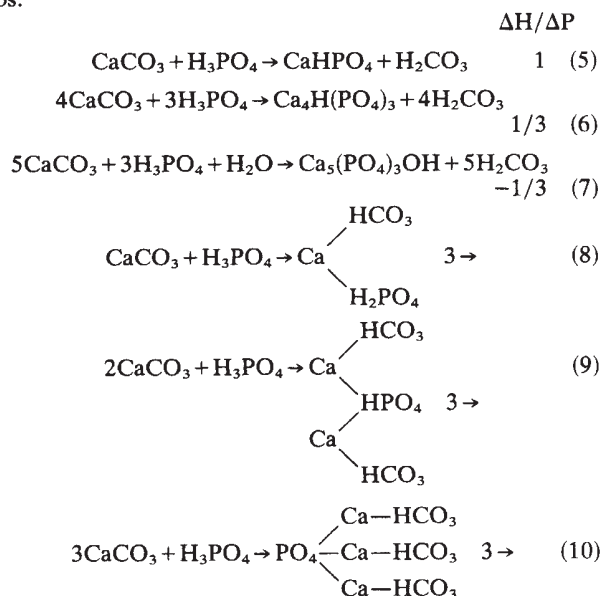
$$\text{Ca (P - p.t.d.)} = -(\Delta\text{Ca} + \Delta\text{CO}_3) \quad (4)$$

where ΔCa and ΔCO_3 are the changes in total calcium and carbonate respectively, from the time of phosphorus addition until the end of the experiment.

The ratio of precipitated Ca to precipitated P, averaged for

the data presented in Table 1, is 0.620. This ratio is not within that defined by equation (3) and thus is not consistent with the hypothesis that a known calcium phosphate shall have precipitated.

Equations (5)–(10) describe six possible reactions between CaCO_3 and phosphoric acid, together with the resultant $\Delta\text{H}/\Delta\text{P}$ ratios.



Reactions (5)–(7), leading to the formation of distinct calcium phosphates (DCP, octa-calcium phosphate or hydroxyapatite) do not lead to an appreciable binding of hydrogen to the solid phase. Reactions (8)–(10), describing the formation of calcium-carbonate-phosphates, lead to an equivalent binding of phosphate and hydrogen ions to the solid phase. The necessary $\Delta\text{H}/\Delta\text{P}$ ratio is 3.

The linear relation found between the transfer of hydrogen and phosphorus from the solution to the solid (equation (2)), yields a $\Delta\text{H}/\Delta\text{P}$ ratio of 2.6. This ratio tends to decrease, however, with the increase in phosphate concentration. This trend, and the fact that the experimental $\Delta\text{H}/\Delta\text{P}$ ratio is lower than that demanded by equations (8)–(10), is probably because some fraction of the phosphate had reacted with calcium according to reactions (5)–(7).

These considerations and results suggest that the reaction between CaCO_3 and orthophosphate can be formulated as



According to equation (11), and the potential diagram suggested by MacGregor and Brown¹⁰, the chemical potentials of the components participating in reactions (8)–(10) can be related through equation (12).

$$n\mu\text{CaCO}_3 + \mu\text{H}_3\text{PO}_4 = \text{F}^\circ \text{ solid} \quad (12)$$

or

$$n \ln(\text{CaCO}_3) + \ln(\text{H}_3\text{PO}_4) = \text{constant} \quad (13)$$

Table 1 Material balance in CaCO_3 suspensions enriched with phosphate

Changes in solution composition (mol l^{-1})				
Initial phosphorus	ΔP	ΔH	ΔCa	ΔCO_3
$2 \cdot 10^{-5}$	$-(3.94 \pm 0.11) \times 10^{-6}$	$-(2.95 \pm 0.23) \times 10^{-5}$	$-(2.32 \pm 0.31) \times 10^{-5}$	$-(2.31 \pm 0.26) \times 10^{-5}$
$4 \cdot 10^{-5}$	$-(7.89 \pm 0.22) \times 10^{-6}$	$-(3.43 \pm 0.28) \times 10^{-5}$	$-(2.49 \pm 0.13) \times 10^{-5}$	$-(1.97 \pm 0.28) \times 10^{-5}$
$8 \cdot 10^{-5}$	$-(1.59 \pm 0.01) \times 10^{-5}$	$-(4.63 \pm 0.26) \times 10^{-5}$	$-(2.38 \pm 0.2) \times 10^{-5}$	$-(1.56 \pm 0.25) \times 10^{-5}$
$1.6 \cdot 10^{-4}$	$-(3.02 \pm 0.04) \times 10^{-5}$	$-(6.17 \pm 0.33) \times 10^{-5}$	$-(1.97 \pm 0.17) \times 10^{-5}$	$-(2.71 \pm 2.02) \times 10^{-6}$
$1.2 \cdot 10^{-4}$	$-(2.37 \pm 0.004) \times 10^{-5}$	$-(3.93 \pm 0.52) \times 10^{-5}$	$-(2.7 \pm 1.95) \times 10^{-6}$	$(4.7 \pm 5.1) \times 10^{-6}$
$2.4 \cdot 10^{-4}$	$-(4.36 \pm 0.03) \times 10^{-5}$	$-(7.65 \pm 0.39) \times 10^{-5}$	$-(5.46 \pm 0.46) \times 10^{-6}$	$(1.18 \pm 0.38) \times 10^{-5}$
$4.8 \cdot 10^{-4}$	$-(7.64 \pm 0.13) \times 10^{-5}$	$-(1.29 \pm 0.03) \times 10^{-4}$	$-(5.81 \pm 0.77) \times 10^{-6}$	$(4.11 \pm 0.34) \times 10^{-5}$
$9.6 \cdot 10^{-4}$	$-(1.15 \pm 0.1) \times 10^{-4}$	$-(2.78 \pm 0.04) \times 10^{-4}$	$-(5.89 \pm 1.19) \times 10^{-6}$	$(2.55 \pm 1.01) \times 10^{-5}$

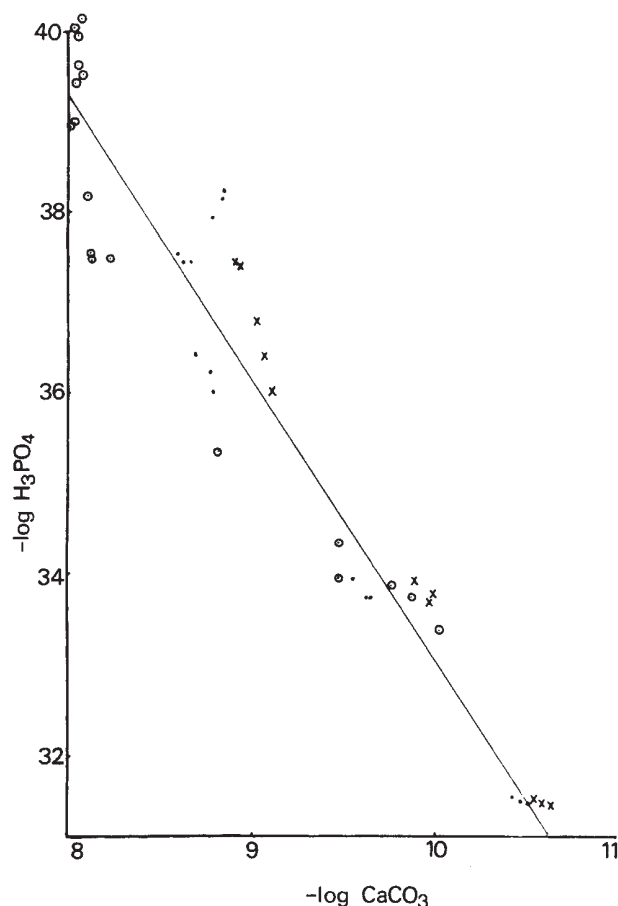


Fig. 1 H_3PO_4 potential as a function of CaCO_3 potential in CaCO_3 suspensions equilibrated with phosphorus adsorbed on anion exchange resin. ●, Experiment (1), 5 g CaCO_3 per 200 ml; ×, experiment (2), 5 g CaCO_3 per 200 ml; ○, experiment (3), 1 g CaCO_3 per 200 ml.

where the values in parentheses represent the activity of the given components.

According to equation (13), the logarithm of CaCO_3 activity plotted against that of H_3PO_4 should yield a straight line, the slope of which is equal to n , the stoichiometric ratio of CaCO_3 to H_3PO_4 .

Calcium carbonate suspensions were equilibrated during three days with phosphated anion exchange resin. The experimentally derived data are presented, in a potential diagram form, in Fig. 1. The points follow a straight line defined by

$$\log (\text{H}_3\text{PO}_4) = 64.23 - 3.11 \log (\text{CaCO}_3); \quad r = -0.951 \quad (14)$$

This line is in accordance with equation (13), and the slope suggests that the reaction occurring and controlling the system is that described by equation (10).

The formation of a surface compound $\text{Ca}_3(\text{HCO}_3)_3\text{PO}_4$ agrees with the crystal structure of calcite or aragonite. In both crystals, three calcium ions are located on the surface for each unit cell¹¹.

The linear relation between the CaCO_3 and H_3PO_4 potentials seems to disagree with the reported relationship between the chemical potentials of $\text{Ca}(\text{OH})_2$ and H_3PO_4 (refs 1, 3), but the present experimental data also yield a straight line when $\log \text{H}_3\text{PO}_4$ is plotted against $\log \text{Ca}(\text{OH})_2$. This ambiguity stems from the fact that in any system at which H_2CO_3 potential is constant (systems in equilibrium with the atmosphere), $\text{Ca}(\text{OH})_2$ potential is related to CaCO_3 potential. Thus, in any system controlled by equation (14), and maintaining a constant $p\text{CO}_2$,

the phosphoric acid potential will show an apparent dependence on $\text{Ca}(\text{OH})_2$ potential.

The existence of the surface compound, $\text{Ca}_3(\text{HCO}_3)_3\text{PO}_4$, also has an effect on the solubility of CaCO_3 . The activity of CaCO_3 in the solution is equivalent to the ionic product $(\text{Ca}) \cdot (\text{CO}_3)$. The data in Fig. 1 clearly indicate that this product is a function of the phosphoric acid potential and not a constant value. When the phosphoric acid potential is high, the ionic product drops significantly below the solubility product of calcite ($pK = 8.31$ (ref. 8)).

This work has been supported by a research grant from the Lake Kinneret Authority.

Received 9 May; accepted 24 September 1980.

1. Cole, C. V. & Olsen, S. R. *Proc. Soil Sci. Soc. Am.* **23**, 116–118 (1959).
2. Gregory, T. M., Moreno, E. C. & Brown, W. E. *J. Res. natn. Bur. Standards* **74A**, 461–475 (1970).
3. Avnimelech, Y. *Verh. int. Verein. Limnol.* **19**, 2305–2308 (1975).
4. Stumm, W. & Leckie, J. O. *Proc. 5th int. Conf. on Advances in Water Pollution Research* (Pergamon, Oxford, 1971).
5. Rootare, H. M., Deitz, V. R. & Carpenter, F. G. *J. Colloid Sci.* **17**, 179–206 (1962).
6. Olsen, S. R. & Watanabe, J. S. *Proc. Soil Sci. Soc. Am.* **21**, 144–149 (1975).
7. Rand, M. C., Greenberg, A. E. & Taras, M. T. (eds) *Standard Methods for the Examination of Water and Wastewater* 14th edn (American Public Health Association, Washington DC, 1976).
8. Nakayama, F. S. *Soil Sci.* **106**, 429–434 (1968).
9. Gimblett, F. G. R. & Mohk, C. B. *Trans. Faraday Soc.* **50**, 965–972 (1959).
10. MacGregor, J. & Brown, W. E. *Nature* **205**, 359–361 (1965).
11. Grimshaw, R. W. *The Chemistry and Physics of Clays and Other Ceramic Materials* (Benn, London, 1971).
12. Avnimelech, Y., Moreno, E. C. & Brown, W. E. *J. Res. natn. Bur. Standards* **77A**, 149–155 (1973).

Plant-controlled supratidal anhydrite from Al-Khiran, Kuwait

A. Gunatilaka, A. Saleh & A. Al-Temeemi

Department of Geology, Kuwait University, Kuwait

We report here that in the supratidal zone of the Al-Khiran sabkha, diagenetic anhydrite occurs exclusively in association with the living halophyte *Halocnemum strobilaceum* (Pallas) M.B. This anhydrite has a hair cream-like consistency and the bulk of it has formed as pseudomorphs after gypsum crystals. On the death of the plant at the end of its life cycle, the anhydrite is hydrated back to gypsum. This two-way transformation results in a series of characteristic textures, whose preservation could be due to the comparative 'youth' of the sabkha, which is in the very earliest stages of evaporite diagenesis.

A recent carbonate-algal stromatolite-evaporite cycle <2-m thick and covering an area >60 km² occurs in the Al-Khiran area of Kuwait. The depositional area is in contact with the open waters of the Arabian Gulf by means of two large tidal channels (khors) which provide the main hydraulic energy input into the system. The physiography and sedimentary pattern of the region is shown in Fig. 1.

The supratidal zone proper or sabkha is marked by the occurrence of the halophyte *Halocnemum strobilaceum* (Pall.) M.B., which covers an area 1–3 km wide and extends continuously for ~15 km. Here the groundwater table is no more than 50–70 cm from the surface. Tidal flooding of this zone has not been observed, but probably occurs when strong winds and high tides coincide. The summer temperatures on the sabkha surface can vary between 42 and 55 °C (April to October) and winter temperatures from 20 to 35 °C (November to March). Evaporation rates are very high and rainfall (1979–80) can be up to 12 cm yr⁻¹. Relative humidity values are generally <50%. The tidal range in the khors is >2 m at the mouth and <30 cm at the extremities. In all cases, the evaporite minerals present were confirmed by X-ray diffraction studies.

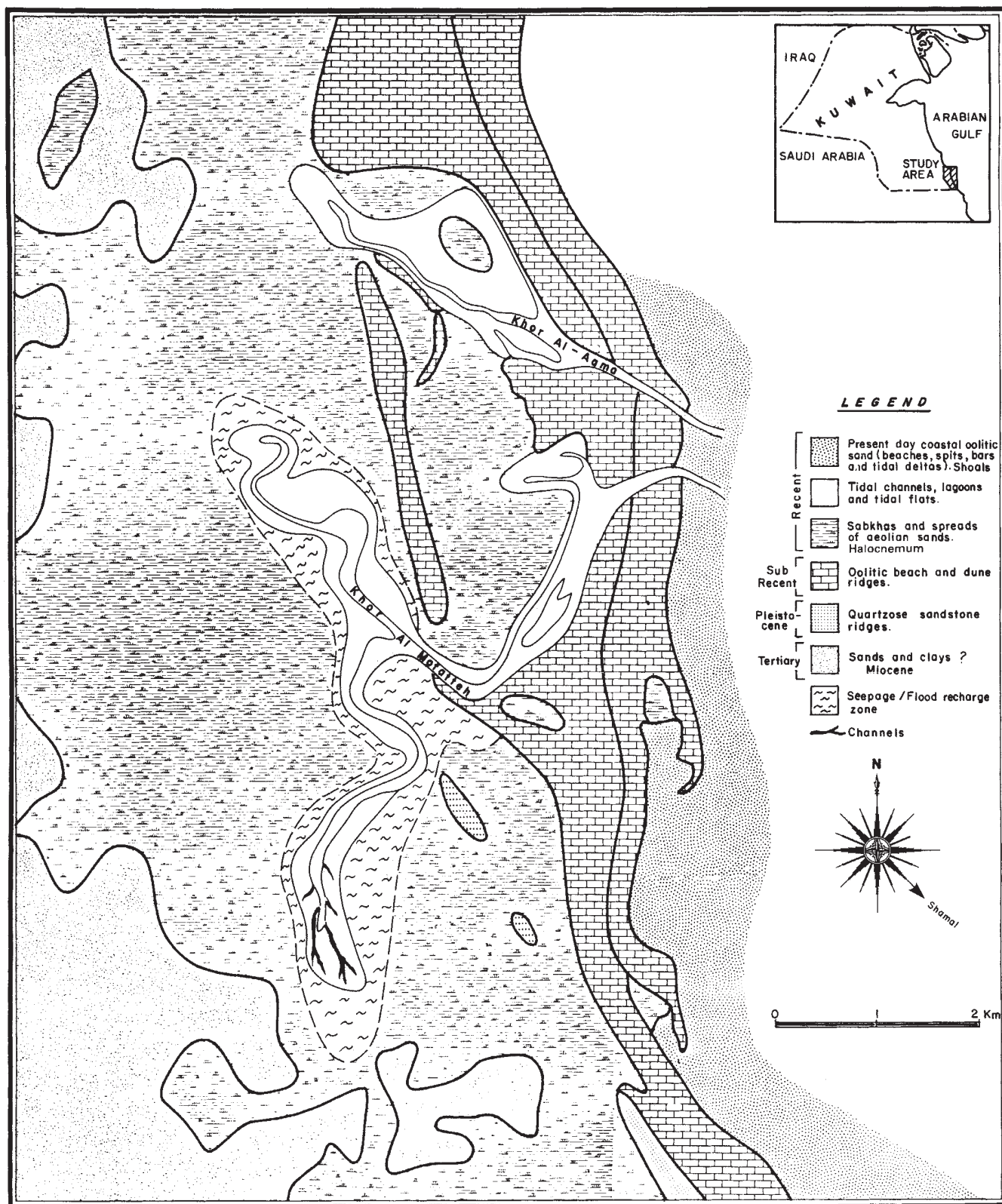


Fig. 1 Physiography and sedimentary pattern of the Al-Khiran area.

The first occurrence of anhydrite in the supratidal zone coincides with the appearance of the *Halocnemum* plant. It occurs as white patches exclusively around the plants, forming small elevated humps above the general surface (Fig. 2). It was also observed that the anhydrite is formed only around the mature, fully grown plants, in layers 3–12-cm thick with the characteristic consistency of hair cream (not the 'cottage cheese'

consistency as in Abu Dhabi). The stratigraphic relationships in a 30-cm core taken through the plant has the features shown in Table 1.

The anhydrite first begins to form as single nodules in the gypsum mush, a few centimetres below the sabkha surface, around the roots. Several of these nodules eventually coalesce to form irregular laminae, which in turn form the anhydrite layers

Table 1 Stratigraphic features of a *Halocnemum* plant taken from the Al-Khiran area of Kuwait

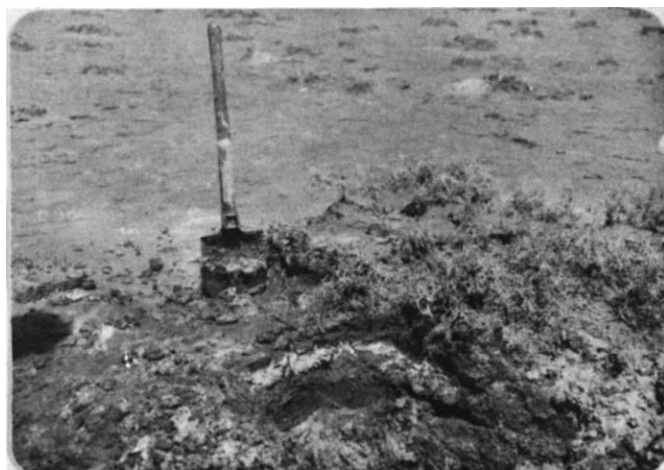
Depth (cm)	Dominant minerals
0–1	Wind-blown material, aragonite, gypsum, anhydrite, dolomite, halite
1–7	Anhydrite nodules or layers, some gypsum + trace dolomite
7–9	Cemented gypsum crust
9–10	Gypsum + some anhydrite
10–19	Gypsum mush + minor aragonite + dolomite
19–25	Aragonite + gypsum + some dolomite
>25	Aragonite mud

around the plant. The continued growth of the gypsum below this zone contorts the layering, eventually resulting in structures resembling enterolithic veins. In section, the anhydrite layer looks like a 'diapir' with a *Halocnemum* bush in the middle. In fact, the surface expression of the anhydrite is directly related to the extent of the root network at depth. The diameter of these patches vary from 20 to 70 cm.

In thin section, the anhydrite occurs as perfect pseudomorphs after diskoid gypsum crystals of variable size (Fig. 3). The pseudomorphs are not formed by the infilling of diskoid voids left by the dissolution of gypsum crystals as suggested by Butler¹. The gypsum–anhydrite transformation is a direct dehydration *in situ*, and the bulk of the nodules in the sabkha formed originally as pseudomorphs after gypsum within the host sediment. The evaporite minerals developing in the sabkha include aragonite, gypsum, bassanite, anhydrite, halite, protodolomite, celestite and magnesite. Detailed microscopic and X-ray diffraction studies indicate that the anhydrite and protodolomite are formed through replacement reactions between gypsum and aragonite and coexisting sabkha brines within the sediment. Note that in all the textures studied, the anhydrite had a gypsum precursor. There is no evidence for any orthochemical precipitation of anhydrite as a primary mineral in this sabkha^{2,3}.

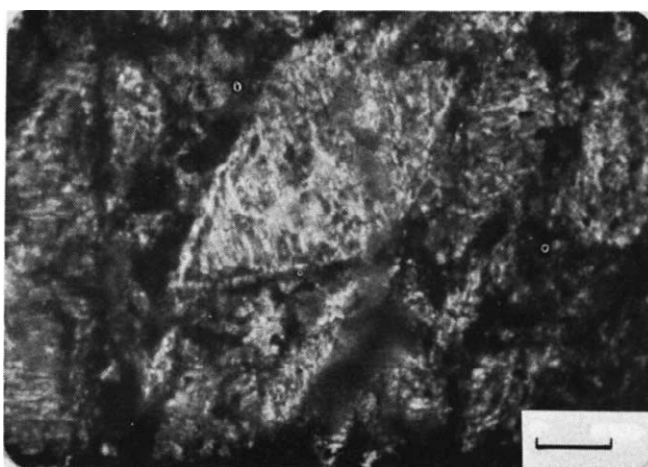
We believe that a brine concentration of 7–8 × SSW (standard seawater of chlorinity 19.377‰) necessary to form the anhydrite, is achieved by combining the effects of evapotranspiration activity of the *Halocnemum*, the osmotic activity of the roots and the cemented gypsum layer below, which regulates the capillary flow of the water. The mass dehydration around the plant results in a series of evaporitic textures caused mainly by the creation of new voids. The gypsum–anhydrite transformation results in a volume reduction⁴ of ~36%.

Many dead plants occur within the supratidal zone. When associated with such plants, the anhydrite changes to a dirty grey or buff colour and the area around the plant becomes hum-

**Fig. 2** Occurrence of hair-cream anhydrite (white patches) around the *Halocnemum* plants.

mocky. Vuggy features tend to develop and the entire surface starts breaking up, by which time the thixotropic consistency of the anhydrite has changed to one resembling moist and hardened flour. Microscopically this change is reflected in the rehydration of the anhydrite back to gypsum. The original pseudomorphs are destroyed and are replaced by a new series of hydration textures which are quite unlike that of any reported before from sabkhas. Around the dead plants optimal physico-chemical conditions probably do not prevail to concentrate the brines sufficiently (7–8 × SSW) and keep the system within the metastable field of anhydrite.

In thin section, the first change observed is the splaying open of the anhydrite pseudomorphs, brought about by the increase in volume due to rehydration (a 66% volume increase)⁴. Most pseudomorphs are flattened along their *c*-axes and the best developed cleavages are normal to the long axis of a crystal. The crystal is split along the latter cleavages, resulting in small, discrete sub-crystals or cleavage laths of anhydrite having rectangular cross-sections. These are very soon hydrated to gypsum laths pseudomorphic after the anhydrite cleavage laths. This is alabastrine secondary gypsum because of its microcrystalline–cryptocrystalline nature. Clumps or aggregates of these laths always show undulating extinction shadows when the microscope stage is rotated. The range of new textures developed vary from collapsing nodules of anhydrite/secondary gypsum laths (both of which originated in diskoid anhydrite

**Fig. 3** Photomicrograph of anhydrite pseudomorph after gypsum from the supratidal *Halocnemum* zone. Scale bar, 0.1 mm.

pseudomorphs) to 'granoblastic and preferred orientation textures'⁵. Perhaps the most interesting intermediate stage is the generation of a nested or 'augen'-like texture, where a nodule of anhydrite pseudomorphs or a single anhydrite grain is surrounded by gypsum laths showing a distinct lineation and indicating some flowage (Fig. 4). Compaction settling and pore water pressures causing liquefaction in poorly draining sediments are believed to cause lineation fabrics in anhydrite laminates⁶. If so, the inter- and intra-nodular permeability characteristics within the sediments could play a crucial part in their genesis. The compact and hardened nature of the alabastrine gypsum is probably due to the decrease in the void space between the laths, caused by the volume increase during hydration.

The hydration mechanism is thought to be dominantly one of dissolution of anhydrite followed by reprecipitation as gypsum⁷. However, there is no evidence that this occurs in Al-Khiran. Finally, when camel herds trek across the *Halocnemum* zone, their footprints soon convert the alabastrine gypsum pseudomorphic laths back into anhydrite, but with an aphanitic, felted texture⁸. Although the latter is probably a result of sustained compaction, followed by slippage or gliding of the breaking microcrystalline laths, the exact mechanisms involved are

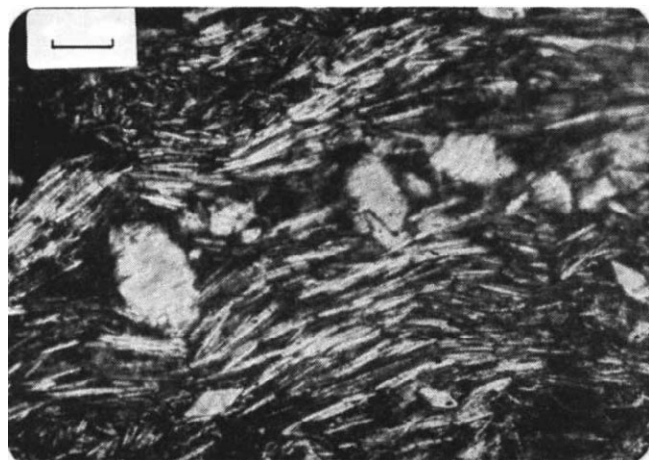


Fig. 4 Photomicrograph showing aligned fabric of gypsum laths. Note the nested anhydrite crystals. Scale bar, 500 μm .

obscure. The main changes observed are:

Diskoid primary gypsum $\rightarrow$
 anhydrite pseudomorphs $\rightarrow$
 anhydrite cleavage laths $\rightarrow$
 alabastrine secondary gypsum pseudomorphic laths $\rightarrow$
 felted-aphanitic anhydrite.

Supratidal anhydrite associated with plants is of interest and there is experimental evidence regarding anhydrite precipitation in the presence of organic matter⁸. The occurrence of anhydrite in Al-Khiran is obviously controlled by the physiological requirements of the halophyte. Whether this difference with the widespread Trucial Coast anhydrite is due to the different climatic regimes (less rainfall and no vegetative cover in Abu Dhabi) is not clear. The absence of anhydrite in the North African sabkhas is probably due to the large influx of continental groundwaters and the semi-arid climatic regime^{9,10}. We have also observed incipient anhydrite formation in a very restricted intertidal channel area, where a reducing organic-rich (algal/bacterial) laminated sediment is forming. Isolated gypsum grains altering to anhydrite were observed 4–5 cm below the intertidal surface in scattered locations. Though quantitatively insignificant, this, albeit rare, occurrence indicates that Perthuisot's oxidizing environment theory for the Arabian Gulf anhydrites needs to be treated with caution¹¹. The Al-Khiran chemical data are yet incomplete to assess the general applicability of this idea.

The present conclusions made regarding the genesis of this unique occurrence of anhydrite in an Arabian Gulf sabkha are tentative because studies are continuing to relate the evaporite mineral assemblages to the chemistry of the coexisting brines within each zone. Many of the features reported here are either absent or not well represented in the classic Trucial Coast sabkhas. The comparatively youthful Al-Khiran sabkhas are still in the 'early' stages of a first cycle of carbonate sedimentation. Here, the intertidal algal mat does not extend inland below the present sabkha surface, and typical chicken-wire anhydrite is absent.

We thank Dr Graham Evans for visiting the study area and making critical comments on many aspects of this study.

Received 6 June; accepted 18 September 1980.

- Butler, G. P. *Ohio geol. Soc.* **1**, 120–152 (1970).
- Kinsman, D. J. J. *Am. Petrol. Inst. Prog. Rep.* (1969).
- Shearman, D. J. *Inst. Min. Metall. Trans.* **B75**, 208–215 (1969).
- Mossop, G. D. & Shearman, D. J. *Inst. Min. Metall. Trans.* **B**, 147–154 (1973).
- Holliday, D. W. *Inst. Min. Metall. Trans.* **B**, 81–84 (1973).
- Shearman, D. J. & Fuller, J. G. *Bull. Can. petrol. Geol.* **17**, 496–525 (1969).
- Ostroff, A. G. *Geochim. cosmochim. Acta* **28**, 1363–1372 (1964).
- Cody, R. D. & Hull, A. B. *Geology* (in the press).
- Rouse, J. E. & Sherif, N. *Nature* **285**, 470–472 (1980).
- West, I. M., Ali, Y. A. & Hilmy, M. E. *Geology* **7**, 354–358 (1979).
- Perthuisot, J. P. *Bull. geol. Soc. Fr.* **19**, 1145–1149 (1977).

Particulate organic carbon flux in the oceans—surface productivity and oxygen utilization

Erwin Suess

School of Oceanography, Oregon State University, Corvallis, Oregon 97331

Organic detritus passing from the sea surface through the water column to the sea floor controls nutrient regeneration, fuels benthic life and affects burial of organic carbon in the sediment record^{1–3}. Particle trap systems have enabled the first quantification of this important process. The results suggest that the dominant mechanism of vertical transport is by rapid settling of rare large particles, most likely of faecal pellets or marine snow of the order of $>200 \mu\text{m}$ in diameter, whereas the more frequent small particles have an insignificant role in vertical mass flux^{4–6}. The ultimate source of organic detritus is biological production in surface waters of the oceans. I determine here an empirical relationship that predicts organic carbon flux at any depth in the oceans below the base of the euphotic zone as a function of the mean net primary production rate at the surface and depth-dependent consumption. Such a relationship aids in estimating rates of decay of organic matter in the water column, benthic and water column respiration of oxygen in the deep sea and burial of organic carbon in the sediment record.

The relationship is based on flux data compiled from particle traps installed throughout the world's oceans^{7–19} by many investigators and from annual mean organic carbon production rates of the respective surface waters^{20–29} (Table 1). In a few cases biological production rates were determined at the time of installing the sediment trap^{12,16}, the remaining data were compiled from other sources, often using annual mean rates from the general area of investigation. The predicted flux may overestimate the proportion of surface production intercepted at any given depth, when traps were installed during times of peak productivity (which was mostly the case) but then were normalized to lower mean annual production rates. Accordingly, equation (1) gives an upper limit of the net vertical carbon flux, if the production is known:

$$C_{\text{flux}(z)} = \frac{C_{\text{prod}}}{0.0238z + 0.212};$$

$$n = 33, \quad r^2 \approx 0.79, \quad z \geq 50 \quad (1)$$

C_{prod} is the primary production rate of carbon at the surface, C_{flux} is the organic carbon flux at depth measured by the sediment traps, both in units of $(\text{g m}^{-2} \text{yr}^{-1})$, $z (\geq 50)$ is the water depth (m) at which the traps were moored, n is the number of observations, and r^2 the coefficient of correlation. Only flux data from below the euphotic zone ($>50 \text{ m}$) are used in defining equation (1). This limitation also causes the non-zero intercept for $b = 0.212$. Moreover carbon fluxes in shallow waters, as in the western Baltic Sea^{12,13} and along cross-shelf transects of the Peru coastal upwelling regime^{16–18}, are consistently less than predicted by extrapolating equation (1) to such depths (Fig. 1). These low yields are probably due to non-vertical motions of particles caused by advection, horizontal currents and locomotion of live plankton, thus exporting particulate matter from the site of sediment trap deployments.

Effects of $\pm 15\%$ uncertainty in primary production estimates are also shown in Fig. 1. This uncertainty is of the same magnitude as the variations and ranges of repeated flux measurements obtained using several traps at the same location. For example, particle flux measurements below the California Current¹⁰ were repeated within the same season whereas the data from the Peru coastal upwelling system^{17,18} are averaged

over several seasons, as indicated in Table 1. The collection and closure efficiency of certain trap designs, the problems of *in situ* decay of non-poisoned collectors, horizontal and resuspended flux probably contribute even more significantly to the scatter of the data^{30,31}.

However large the present uncertainties may be for the relationship of vertical fluxes of particulate carbon in the oceans, they have interesting implications. First, if depth can be adequately transformed into time, the slope of equation (1) implies some sort of rate constant. Actual measurements of large particle size spectra with depth are not available. For small particles ($d < 100 \mu\text{m}$) many theoretical models and empirical distribution patterns have been presented which suggested that particle populations get smaller in diameter with increasing depth^{4,32}. The decrease in particle diameter with depth is thought to be a function of various chemical dissolution or oxidation processes³³⁻³⁵. Thus one would expect a gradual decrease in particle settling rates with depth. However, recent investigations suggest that biological repacking by successive feeding actually produces larger particles at the expense of smaller ones and that settling rates are extremely rapid³⁶. A relationship of carbon-flux to surface-production from which minimum sinking rates can be inferred was recently also observed by Deuser and Ross¹⁴ who showed that seasonality in surface production is reflected in like changes of the carbon flux in the deep Sargasso Sea. The coupling between production at the surface and deep flux—here at a depth of 3,200 m—within a period of a month or less, requires sinking rates of particles in excess of 100 m day^{-1} . Also Shanks and Trent⁶ measured sinking rates of $\sim 50\text{--}100 \text{ m day}^{-1}$ for other large aggregates (marine snow) primarily involved in vertical mass transfer.

As a first approximation, then, depth and residence time may be linearly related and for simplicity a settling velocity of $w_s = 100 \text{ m day}^{-1}$ may be assumed in transforming depth (z) of equation (1) into time. This then enables the definition of an

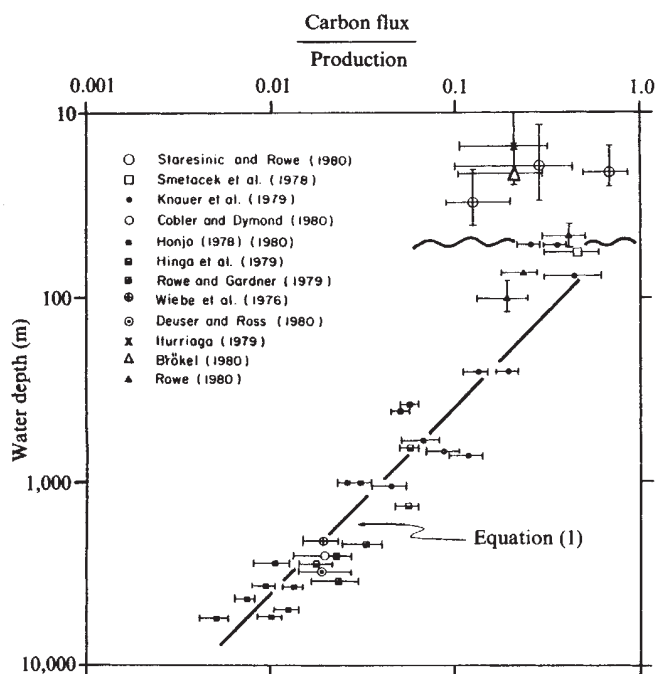


Fig. 1 Organic carbon fluxes with depth in the water column normalized to mean annual primary production rates at the sites of sediment trap deployment. The undulating line indicates the base of the euphotic zone; the horizontal error bars reflect variations in mean annual productivity as well as in replicate flux measurements during the same season or over several seasons; vertical error bars are depth ranges of several sediment trap deployments and uncertainties in the exact depth location. The data points by Rowe (1980) represent selected averages of 2–5 single sites at $\sim 10 \text{ m}$ above the bottom, where resuspension was assumed to be minimal.

Table 1 Organic carbon fluxes in the water column and mean primary production rates

Latitude	Longitude	Water depth (m)	Carbon flux ($\text{g m}^{-2} \text{ yr}^{-1}$)	Production ($\text{g m}^{-2} \text{ yr}^{-1}$)	Refs
31°32.5' N	55°55.4' W	976	0.89	40	8, 29
		3,694	0.32		
		5,206	0.026(?)		
31°32.0' N	55°00.8' W	5,367	0.45	45	8, 29
		389	2.46	50	
		988	1.44		
15°21.1' N	151°28.5' W	3,755	0.63		8, 26
		5,068	0.62		
		378	1.30	40	
25 nm SE Bermuda	0°36.1' N 86°6.1' W	978	0.20(?)		14, 23
		2,778	0.40	100	
		4,280	0.32	15, 22	
27°42.1' N	78°54.1' W	5,582	0.24		9, 24, 26
		3,200	0.77	40	
		2,570	2.0	100	
33°30.1' N	76°15.1' W	660	5.3	100	7
		1,350	10.9	200	
		38°23.1' N	5.6	250	
24°51.1' N	77°39.1' W	2,100	2.0 $\pm$ 0.3	100	10, 20, 21
		36°42.1' N	158 $\pm$ 5	500	
		250	92 $\pm$ 5		
36°42.1' N	122°13' W	700	42 $\pm$ 2		150
		50	33 $\pm$ 7		
		250	19 $\pm$ 2		
32°47.1' N	144°26' W	700	17 $\pm$ 2		100
		75	25 $\pm$ 10		
		575	5.3 $\pm$ 1		
38°50.1' N	72°31.1' W	1,050	4.4 $\pm$ 2		11, 26, 28
		2,160	6.4 $\pm$ 0.2	200	
		38°28.5' N	3.5 $\pm$ 0.7	200	
38°19.1' N	69°37.1' W	3,500	5.4 $\pm$ 1.0	250	12
		55°38.6' N	65	460 $\pm$ 150	
		7°39.5' S	378 $\pm$	650	
10°04.2' S	78°11.2' W	22	370 $\pm$	450	17, 25
		11°58.5' S	253 $\pm$	300	
		13°40.4' S	167 $\pm$	350	
15°04.5' S	75°26.6' W	50 $\pm$ 10	240 $\pm$ 100 $\pm$	1,200 $\pm$ 600	16, 17, 25
		15°05.4' S	130 $\pm$ 20 $\pm$	600	
		15°06.3' S	110 $\pm$ 30 $\pm$	600	
53°30.1' N	10°02.1' E	100 $\pm$ 10	30 $\pm$ 10	150	13, 19, 27
		< 20			

* Plotted as one datum point.

† Flux measured during one season only.

‡ Flux measured during two or more seasons.

§ Productivity determined during the same season when sediment traps deployed.

apparent rate constant for organic carbon decomposition from equation (2):

$$\frac{1}{C_{\text{flux}(z)}} - 0.212 = 2.38t \quad (2)$$

where C_{prod} = production at $t = 0$, and the constant $k = 2.38$, (that is $z(0.0238)$), has the dimension day^{-1} and the properties of a second-order rate constant. This is inferred because equation (2) is analogous to the expression for second-order rates in chemical kinetics: $1/C_t - 1/C_0 = kt$. In testing experimental data to ascertain second-order rate kinetics, $1/C_t$ plotted against t should yield a straight line relationship whose slope is equal to the rate constant k and the intercept equal to $1/C_0$. Accordingly, a fit for the carbon flux data in relation to depth (z), to settling velocity (w_s) and to primary production C_{prod} defines the equation above.

As a second implication, the extrapolation of the predicted carbon flux from equation (1) to the interception with the sea floor at any site yields the net carbon sedimentation before burial. This is independent of the actual rate laws and mechanisms that control organic matter consumption at depth or the variation in sinking rates of particulate matter. Net carbon sedimentation is an important parameter because it gives the upper limits of organic carbon accumulation in the sediment column. The preserved rate of burial can only be smaller than this flux. The magnitude of continued consumption of organic matter from the time it arrives at the sediment surface to the

Table 2 Organic carbon utilization by benthic processes at two sites of the coastal upwelling regime off Peru

	Station 7706-39	Station 7706-36
Latitude	11°15.1' S	13°37.3' S
Longitude	77°57.4' W	76°50.5' W
Water depth (m)	186	370
C-org (g m ⁻² yr ⁻¹)		
Production annual mean	500	400
Flux to seafloor using equation (1)	108	44
Burial rate from sediment accumulation ³⁷	40	17
Benthic utilization as carbon difference ³⁷	68	38*–142†
O ₂ (l m ⁻² yr ⁻¹)		
Oxygen demand	160	93–345†
		63

* Equivalent oxygen consumption rates are based on 2.43 ml O₂ consumed = 1.0 mg C utilized⁴¹.

† Ref. 38.

time it is effectively removed from decomposition through burial is related to benthic respiration. In many cases the burial rate of organic carbon can be independently determined from linear bulk sedimentation rates, sediment densities, porosities and organic carbon contents. Consequently, the benthic respiration is the difference between the net vertical flux as predicted by equation (1) and the burial rate. The burial rate of organic matter, in turn, seems to be primarily controlled by the linear rate of sedimentation, as discussed elsewhere³⁷. Results in Table 2 at two stations from the coastal upwelling regime off Peru illustrate how to estimate benthic respiration from vertical carbon flux minus carbon burial rate. The carbon utilization rates and the inferred bottom oxygen demand are well within the ranges reported from other coastal upwelling areas. Based on benthic production and faunal analyses, Christensen and

Packard³⁸ estimated benthic oxygen utilization rates at water depths ranging from 20 to 200 m for coastal upwelling areas off north-west Africa and Baja California to be between 93 and 345 (l m⁻² yr⁻¹). Also, oxygen utilization rates of 60–550 (l m⁻² yr⁻¹), with strong seasonal fluctuations, are reported from the first direct measurements of O₂-distribution profiles *in situ* and from incubation experiments of near-shore anoxic sediments from the western Baltic Sea^{39,40}.

The most far-reaching, yet tentative implication, is coupling oxygen consumption in the water column to the observed changes in vertical flux of particulate organic carbon. If for every 106 mol organic carbon consumed, 138 mol oxygen are utilized⁴¹ and if the residence time of the particles within a parcel of seawater can be ascertained, rates of oxygen utilization at any depth can be predicted. Again for simplicity, assuming a mean sinking rate for particles $w_s = 100 \text{ m day}^{-1}$ yields the distribution of the total cumulative oxygen utilization with depth as:

$$\Delta O_2 = C_{\text{prod}}(6.66 \times 10^{-5}) \left[1 - \frac{1}{0.0238(z) + 0.212} \right] \quad (3)$$

where the constant 6.66×10^{-5} converts carbon consumption to oxygen utilization in units of (ml l⁻¹ day⁻¹); that is $\frac{1}{365} \text{ g C-org}$

m⁻² day⁻¹ utilizes $2.43 \times \frac{1}{100} \text{ O}_2 \text{ m}^3$ when assuming settling velocity

of particles $w_s = 100 \text{ m day}^{-1}$; therefore, $\frac{2.43}{365} \times 10^{-5} \times 10^3 =$

$6.66 \times 10^{-5} \text{ ml l}^{-1} \text{ day}^{-1}$. The proportion of organic carbon consumed is given by the expression $1 - \frac{1}{0.0238(z) + 0.212}$ and

again, z is limited to depths $\geq 50 \text{ m}$. To illustrate the usefulness of equation (3), vertical distribution profiles of oxygen consumption rates are generated for a water-column profile where the input of oxidizable particulate organic matter from the euphotic zone is between 100 and 500 g C m⁻² yr⁻¹ and the vertical mass transfer is assumed to be primarily through large particles at sinking rates of 100 m day^{-1} . Comparable oxygen consumption rates in water-column profiles deduced from electron transport activities, bacterial utilization rates and respirometer measurements are in general agreement with these predicted values, although ETS O₂-consumption rates appear higher by an order of magnitude at greater water depths (Fig. 2)^{42–45}. Equation (3) also assumes O₂-consumption to be only by the utilization of particulate organic matter, whereas actually, dissolved organic carbon is also utilized⁴⁶.

The oxygen distribution resulting from the predicted consumption rates of equation (3) largely follows that of the classical circulation model by Wyrtki⁴⁷, the basic difference being that his model utilizes an exponential oxygen consumption term with depth following first-order rate kinetics, that is $C_t = C_0 e^{-kz}$, whereas the carbon flux measurements presented here are better approximated by a second-order fit. The only way that the carbon flux data would follow a first-order rate expression is in the case of accelerated settling of particles; that is, if sinking rates were actually to increase with depth in the water column.

These early results from particle traps apparently tie together successfully processes of organic carbon cycling in the oceans that previously have been treated separately by marine biologists, chemists and sedimentologists. Sizes and shapes of particles involved in vertical mass flux and seasonally adjusted surface productivity and deep carbon fluxes hold the key to understanding fully organic carbon cycling in the ocean and related oxygen and nutrient distributions.

This work was supported by the NSF and the Office of Naval Research through grants OCE-77-20376 and N00014-79-C-0004, Project NRO83-1026 to Oregon State University. I thank J. Dymond, P. D. Komar and R. Pytkowicz for helpful discussions.

I dedicate this work to Richard Cobler, an inspiring student of oceanography and a friend.

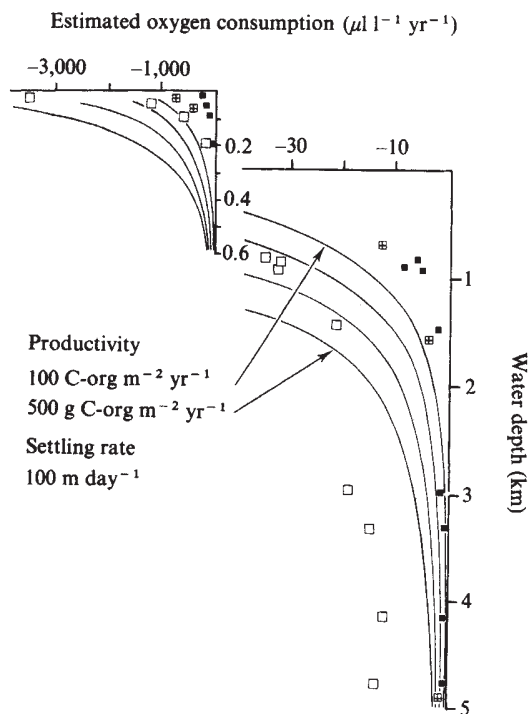


Fig. 2 Vertical distribution of oxygen consumption rates in the ocean derived from changes of particulate carbon fluxes. The profiles are generated from equation (3) for mean primary productivities of between 100 and 500 g C-org m⁻² yr⁻¹, respectively. Vertical transfer of particles is at constant sinking rate of 100 m day^{-1} . Maximum potential oxygen consumption rates are shown for comparison from electron transport activities (□) and the corresponding estimated actual oxygen utilization (■). □, Oxygen consumption from bacterial utilization rates. The data points represent a composite profile from the eastern Pacific Ocean^{42–45}; note scale changes.

Received 18 June; accepted 10 September 1980.

1. Redfield, A. C., Ketchum, B. H. & Richards, F. A. in *The Sea*, Vol. 2 (ed. Hill, M. N.) 26–77 (Wiley, New York, 1963).
2. Smith, K. L. *Mar. Biol.* **47**, 337–347 (1978).
3. Müller, P. J. & Suess, E. *Deep-Sea Res.* **26**, 1347–1362 (1979).
4. McCave, I. N. *Deep-Sea Res.* **22**, 491–502 (1975).
5. Schrader, H.-J. *Science* **174**, 55–57 (1971).
6. Shanks, A. L. & Trent, D. *Deep-Sea Res.* **27**, 137–143 (1980).
7. Wiebe, P. H., Boyd, S. H. & Winget, C. J. *mar. Res.* **34**, 341–354 (1976).
8. Honjo, S. *J. mar. Res.* **36**, 469–492 (1978); **38**, 53–97 (1980).
9. Hinga, K. R., Sieburth, J. & Heath, G. R. *J. mar. Res.* **37**, 557–579 (1979).
10. Knauer, G. A., Martin, J. H. & Bruland, K. W. *Deep-Sea Res.* **26**, 97–108 (1979).
11. Rowe, G. T. & Gardner, W. D. *J. mar. Res.* **37**, 581–600 (1979).
12. Smetacek, V., von Brökel, K. & Zeitschel, Zenk, W. *Mar. Biol.* **47**, 211–226 (1978).
13. Iturriaga, R. *Mar. Biol.* **55**, 157–169 (1979).
14. Deuser, W. G. & Ross, E. H. *Nature* **283**, 364–365 (1980).
15. Cobler, R. & Dymond, J. *Science* **209**, 801–802 (1980).
16. von Brökel, K. *Abstr. IDOE int. Symp. on Coastal Upwelling*, Los Angeles (1980).
17. Rowe, G. T. *Symp. on Bio-productivity of Upwelling Ecosystems*, Moscow (1979).
18. Staresinic, N. & Rowe, G. T. *Abstr. IDOE int. Symp. on Coastal Upwelling*, Los Angeles (1980).
19. Zeitschel, B. *Kieler Meeresforsch.* **21**, 55–80 (1965).
20. Owen, R. W. *California Cooperative Oceanic Fisheries Investigations*, Atlas No. 20, 98–109 (State of California, Marine Research Committee, 1974).
21. Ryther, J. H. *Science* **166**, 72–76 (1969).
22. Love, C. M. & Allen, R. M. (eds) *Eastropac Atlas 10* (US Government Printing Office, Circ. 330, 1975).
23. Menzel, D. W. & Ryther, J. H. *Deep-Sea Res.* **6**, 351–367 (1960).
24. Ryther, J. H. in *The Sea*, Vol. 2 (ed. Hill, M. N.) 347–380. (Wiley, New York, 1963).
25. Guillen, O., de Mendiola, B. R. & de Rondon, R. I. *Oceanography of the South Pacific* (New Zealand National Committee for UNESCO, Publ. 405–418, 1973).
26. Koblenz-Mishke, O. J., Volkovinsky, V. V. & Kabanova, J. G. *Scientific Exploration of the South Pacific*, (ed. Wooster, W.) 183–193. (National Academy of Sciences, Washington, 1970).
27. von Bodungen, B. thesis, Univ. Kiel (1975).
28. Ryther, J. H. & Yentsch, C. S. *Limnol. Oceanogr.* **3**, 327–335 (1958).
29. Ryther, J. H. & Menzel, D. W. *Deep-Sea Res.* **6**, 235–238 (1960).
30. Dymond, J. *et al. Earth planet. Sci. Lett.* (in the press).
31. Kirchner, W. B. *Limnol. Oceanogr.* **20**, 657–660 (1975).
32. Brun-Cottan, J. C. *J. geophys. Res.* **81**, 1601–1606 (1976).
33. Lal, D. & Lerman, A. *J. geophys. Res.* **80**, 423–430 (1975).
34. Wollast, R. *The Sea* Vol. 5 (ed. Goldberg, E. D.) 369–392 (Wiley, New York, 1974).
35. Lal, D. *Science* **198**, 997–1008 (1977).
36. Honjo, S. & Roman, M. R. *J. mar. Res.* **36**, 45–75 (1978).
37. Müller, P. J. & Suess, E. *Deep-Sea Res.* **26**, 1347–1362 (1979); *Colloq. int. du C.N.R.S. No.* **293**, 17–26 (1980).
38. Christensen, J. P. & Packard, T. T. *Deep-Sea Res.* **24**, 331–343 (1977).
39. Revsbeck, N. P., Jørgensen, B. B. & Blackburn, T. H. *Science* **207**, 1355–1356 (1980).
40. Revsbeck, N. P., Sørensen, J., Blackburn, T. H. & Lomholt, J. P. *Limnol. Oceanogr.* **25**, 403–411 (1980).
41. Richards, F. A. *Chem. Oceanogr.* **1**, 611–645 (1965).
42. Packard, T. T., Healy, M. L. & Richards, F. A. *Limnol. Oceanogr.* **16**, 67–70 (1971).
43. Kroopnick, P. *Deep-Sea Res.* **21**, 211–227 (1974).
44. Craig, H. *J. geophys. Res.* **76**, 5078–5086 (1971).
45. Carlucci, A. F. & Williams, P. M. *Naturwissenschaften* **65**, 541–542 (1978).
46. Williams, P. M. & Carlucci, A. F. *Nature* **262**, 810–811 (1976).
47. Wyrski, K. *Deep-Sea Res.* **9**, 11–23 (1962).

Concomitant photoautotrophic growth and nitrogenase activity by cyanobacterium *Plectonema boryanum* in continuous culture

H. W. Pearson & R. Howsley

Department of Botany, University of Liverpool, PO Box 147, Liverpool L69 3BX, UK

Since the discovery that nitrogenase activity will develop in cultures of *Plectonema boryanum* in microaerobic conditions¹, this phenomenon has been demonstrated in numerous other non-heterocystous cyanobacteria^{2–4}. However, there is controversy as to whether such organisms are capable of sustained concomitant nitrogenase activity and photoautotrophic growth^{2,5,6}. An exogenous carbohydrate source (fructose) and strongly reducing conditions (produced by bubbling liquid cultures with a N₂/CO₂/H₂S gas mixture) have been deemed necessary for continuous logarithmic green growth and nitrogenase activity by batch cultures of *Plectonema boryanum* in the light. We report here sustained exponential growth and nitrogenase activity by *P. boryanum* in continuous culture gassed with nitrogen in the absence of further chemical reducing agents and exogenous carbohydrate sources.

An aerobic continuous culture of *P. boryanum* was established in a chemostat with nitrate (1 mM) in the medium as the combined-nitrogen source and light intensity the factor controlling the rate of growth. General conditions for culture were as described in Fig. 1 legend, except that the fermenter was gassed with air (instead of nitrogen) and the rate of dilution was faster ($D = 0.072 \text{ h}^{-1}$ rather than 0.048 h^{-1}). With the population in steady state in these carbon-limited growth conditions, the ratio of phycobiliprotein to chlorophyll *a* (an indication of the nitrogen status of the cells^{1,5}) was 1.030, the cell density 0.115 absorbance units and the cell doubling time 9.6 h. After 10 d of steady-state growth, nitrogen starvation was imposed by omitting nitrate from the inflowing medium. The culture turned yellow within ~36 h, growth virtually stopped (as evidenced by a rapid rate of washout) and the ratio of phycobiliprotein to chlorophyll *a* dropped from 1.030 to 0.903. Thus the higher ratio was obtained when nitrogen was not limiting growth, whereas the lower ratio occurred in conditions of nitrogen starvation.

We used the basic conditions established for aerobic photoautotrophic growth of *P. boryanum* in continuous culture; cells previously grown with nitrate were inoculated into combined-nitrogen-free medium under an atmosphere of nitrogen and grown in the fermenter for 3–4 d in batch culture. During this period nitrogenase activity commenced and cell density increased. The fermenter was then switched to the continuous mode by pumping combined-nitrogen-free medium through the culture while bubbling it with oxygen-free nitrogen gas (inorganic carbon was supplied as bicarbonate in the inflowing medium). The culture stabilized to give a steady-state cell density after approximately 48 h. Figure 1 shows that simultaneous photosynthesis, nitrogenase activity and photoautotrophic exponential growth were sustained by *P. boryanum* in this microaerobic continuous culture. The filaments showed no visible signs of yellowing in spite of a ratio of phycobiliprotein to chlorophyll *a* of 51.56% of the value recorded for the aerobic cultures grown with nitrate.

Our nitrogenase activity (1.00 nmol C₂H₄ per mg protein per min) for *P. boryanum* in continuous culture was substantially less than that quoted for batch cultures grown in a sulphide-sat amended with fructose⁵. Nevertheless it provided sufficient fixed nitrogen to support a higher specific growth rate ($\mu = 0.048 \text{ h}^{-1}$), equivalent to a cell doubling time of 14.4 h compared to ~24 h reported by Rogerson⁵. The growth pattern described here could be maintained almost indefinitely.

We next investigated the effect of adding sulphide at levels known to suppress oxygenic photosynthesis. To establish the range of inhibitory concentrations, we incubated samples of the culture with various concentrations of sulphide: photosynthesis was severely inhibited above 300 μM . We therefore incorporated sulphide into the inflowing medium and it reached a concentration of 300 μM in the fermenter after 20 h, at which time the photosynthetic rate of *P. boryanum* had dropped to 14% of its initial steady-state value. In spite of this severe reduction in carbon fixation, nitrogenase activity was stimulated sevenfold and the specific growth rate of the culture steadily increased for 48 h before falling again in response to carbon starvation. Thus, while previously maintaining a high rate of growth in nitrogen-fixing photoautotrophic conditions, *P. boryanum* accumulated sufficient photosynthetically derived carbon to support a prolonged period of increased growth in the light, in conditions favouring nitrogen fixation but inhibiting photosynthesis. The stimulation of nitrogenase by the added sulphide was reflected in the increased production of phycobiliprotein relative to chlorophyll *a* to a value higher than during growth with nitrate.

Although sulphide reduced the dissolved oxygen concentration in the medium to $31.7 \pm 1.7 \mu\text{M}$, the studies of Stewart and Lex¹ suggest that this alone could not account for all the increase in nitrogenase activity. In microaerobic conditions the controlling factor seems to be intracellular accumulation of photosynthetically produced oxygen, rather than the diffusion of

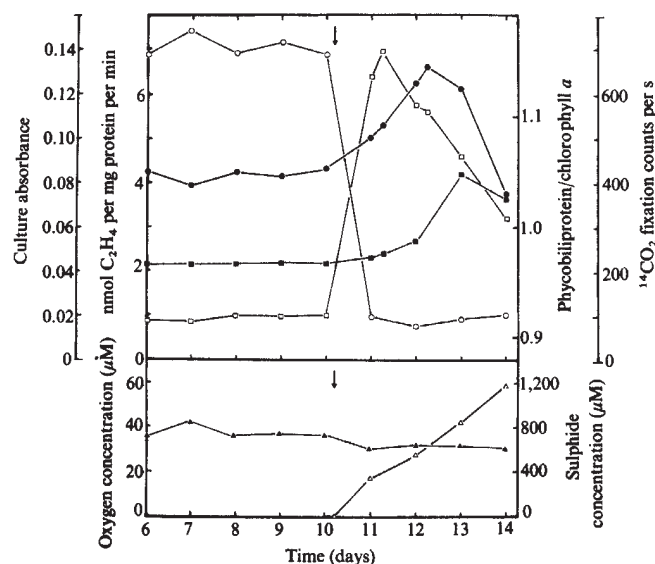


Fig. 1 Nitrogen-fixing photoautotrophic growth of *Plectonema boryanum* (UTEX 594) in chemostat continuous culture. The growth medium was a modified ASM recipe¹¹, free of combined-nitrogen but with NaHCO_3 (12 mM final concentration) added as the source of carbon dioxide. The culture (1,000- cm^3 volume) had a dilution rate of $48 \text{ cm}^3 \text{ h}^{-1}$ and was gassed with oxygen-free nitrogen ($100 \text{ cm}^3 \text{ min}^{-1}$) and stirred continuously. The pH of the medium in the fermenter was adjusted to 8.9. When required, sulphide was pumped into the fermenter (the arrow denotes the start of this procedure) from a separate reservoir of Na_2S solution maintained in anoxic conditions. The dilution rate of the culture was kept constant. Illumination by 125-W white fluorescent tubes provided a light intensity at the fermenter surface of $97.5 \mu\text{E m}^{-2} \text{ s}^{-1}$ photosynthetically active radiation (6,000 lux) and the temperature was maintained at $29 \pm 1^\circ\text{C}$. Dissolved oxygen ($\blacktriangle$) was determined using a Clark-type oxygen electrode housed in a small chamber remote from the main fermenter body into which a sample of the culture could be diverted when measurements were to be made. This prevented poisoning of the electrode by prolonged contact with the medium when it contained sulphide. The optical density of the culture ($\bullet$) was estimated spectrophotometrically at 540 nm (1 cm pathlength) and the phycobiliprotein to chlorophyll *a* ratio ($\blacksquare$) calculated from measurements at 625 and 678 nm respectively. Nitrogenase activity ($\square$) was estimated by the acetylene reduction technique in a gas phase of 10% C_2H_2 balance nitrogen, as described previously¹¹ but by means of a Perkin Elmer FII gas chromatograph. Sulphide concentration ($\triangle$) was determined by the methylene blue photometric method¹². Rates of carbon fixation ($\circ$) were measured as before⁷ except that $\text{NaH}^{14}\text{CO}_3$ was added directly to samples from the fermenter.

oxygen into the cells from the culture medium. Sulphide thus allowed increased expression of nitrogenase activity by penetrating the cells⁷ and reducing the intracellular oxygen tension by two possible mechanisms: (1) the chemical reduction of photosynthetically produced oxygen, as was likely in Rogerson's work⁵, because 5–8 μM sulphide at pH 7.8 does not appreciably inhibit oxygenic photosynthesis in *Plectonema* (Fig. 2), and (2) the direct suppression of photosynthetic oxygen production as in our experiments (Fig. 1).

Nagatani and Haselkorn⁸ have suggested that the shortage of reductant is the major factor limiting induced nitrogenase activity in microaerobic cultures of *Plectonema*. If so the increased rate of nitrogenase activity in the presence of sulphide might be due not just to the elimination of molecular oxygen *in vivo* but also to the supply of reductant directly to the nitrogenase system by sulphide. This could explain the high rates of acetylene reduction we observed with sulphide present, even

when the specific growth rate had started to decline rapidly after day 12 (Fig. 1) due to carbon starvation. In such conditions supplies of reductant and ATP from respiration are likely to be severely restricted. But because sulphide, at the concentrations used, inhibits photosynthetic electron transport through photosystem II while photosystem I is unaffected, cyclic photophosphorylation could provide the ATP in the light and sulphide could donate the bulk of the electrons directly to the nitrogenase system.

Because nitrogenase activity due to stimulation by sulphide leads to the apparent 'luxury' production of phycobiliprotein, as a nitrogenous reserve compound, it seems unwise to draw conclusions about the functioning of photosystem II based solely on a high phycobiliprotein to chlorophyll *a* ratio especially as sulphide inhibited virtually all photosynthetic activity in our culture. Nevertheless, our results point to the potential for renewed photosystem II activity should inhibitory sulphide concentrations prove transitory, as in many natural environments such as temperate salt marshes, mud flats and sand dune slacks.

Our data show that in microaerobic continuous culture *P. boryanum* can sustain a high rate of exponential growth based solely on its own photosynthetic and nitrogen-fixing capacities. The simultaneous and stable rates of photosynthesis and nitrogenase activity do not support the idea of temporal separation of these processes as a mechanism for protecting the nitrogenase from inactivation by oxygen⁹.

Sulphide probably stimulates nitrogenase activity by eliminating intracellular photosynthetically produced oxygen, but sulphide ions may also provide reductant to the nitrogenase system. Oxygenic photosynthesis was markedly more sensitive to sulphide poisoning than was nitrogenase activity but *P. boryanum* can exploit this situation by using stored photosynthate as shown here, or by switching to photoheterotrophic growth if suitable exogenous substrates are available¹⁰.

Given the apparent absence of a special cellular mechanism to permit microaerobic nitrogen-fixing photoautotrophic growth by *P. boryanum*, it is likely that other non-heterocystous cyanobacteria with oxygen-sensitive nitrogenase systems *in vivo* could grow similarly.

R.H. was supported by a NERC studentship.

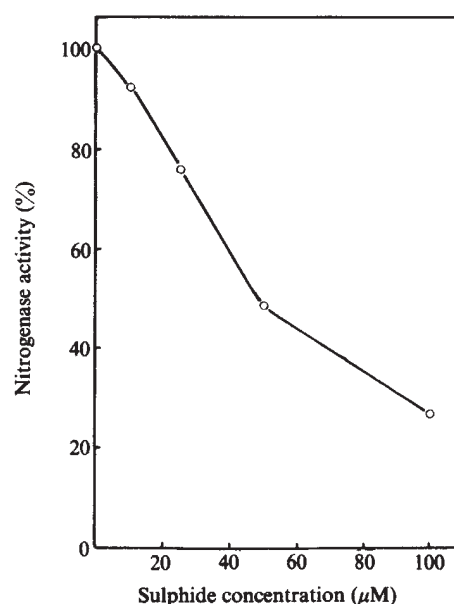


Fig. 2 Carbon fixation by *P. boryanum* at different sulphide concentrations (pH 7.8). Values are expressed as a percentage of the sulphide-free controls and are the means of triplicate samples. Experimental procedure was similar to that described previously⁷.

Received 1 August; accepted September 19, 1980.

1. Stewart, W. D. P. & Lex, M. *Arch. Microbiol.* **73**, 250–260 (1970).
2. Rippka, R. & Waterbury, J. B. *FEMS Microbiol. Lett.* **2**, 83–86 (1977).
3. Kenyon, C. N., Rippka, R. & Stanier, R. Y. *Arch. Microbiol.* **83**, 216–236 (1972).
4. Pearson, H. W. & Tomlinson, D. L. in *Proc. Second int. Symp. Photosynthetic Prokaryotes* (eds Codd, G. A. & Stewart, W. D. P.) 161–163 (University of Dundee, 1976).
5. Rogerson, A. C. *Nature* **284**, 563–564 (1980).
6. Fay, P. & Kulasooriya, S. A. *Br. phycol. J.* **8**, 51–57 (1973).
7. Howsley, R. & Pearson, H. W. *FEMS Microbiol. Lett.* **6**, 287–292 (1979).
8. Nagatani, H. N. & Haselkorn, R. *J. Bact.* **134**, 597–605 (1978).
9. Weare, N. M. & Benemann, J. R. *J. Bact.* **119**, 258–265 (1974).
10. Khoja, T. M. & Whitton, B. A. *Br. phycol. J.* **10**, 139–148 (1975).
11. Stewart, W. D. P. & Pearson, H. W. *Proc. R. Soc., Lond. B175*, 293–311 (1970).
12. Trüper, H. G. & Schegel, H. G. *Antonie v. Leeuwenhoek* **30**, 225–238 (1964).

The robustness of natural systems

Alan Roberts* & Ken Tregonning†

*Physics Department, Monash University, Clayton, Victoria 3168, Australia

†Research Branch, Forests Commission, 1 MacArthur Place, Melbourne, Victoria 3000, Australia

Although the relation between complexity and stability in natural systems has been widely investigated¹, a number of other basic questions remain largely unstudied. For example: can we find fundamental properties which identify a persistent natural system? Are these properties due to the process by which such a system emerges? Is this process—natural selection—also responsible for that robustness which is so generally superior to that of an artificial system? This peculiar robustness of a natural system can show itself in a wide variety of ways; the resistance to shock and efficient nutrient conservation of a climax community², for example, or the physical stability of proteinoid microparticles³. Can we find other system properties which help to explain this robustness—properties which transcend the particular type of system under consideration? Such a 'universal' set of requirements is indeed suggested by the results reported here. In the model studied, the survivor systems turn out to be composed, at every subsystem level, of subsystems each of which is independently viable. These viable subsystems are nested, overlapping, strongly connected and extraordinarily numerous, implying a 'defence in depth' for the main system against major collapse. Some important conclusions can be made about these model systems: first, an artificial system comparable in its size and complexity could hardly have this robust structure built into it; second, if grossly damaged, a natural system will usually be reduced to one of its viable subsystems. The components eliminated could, of course, include ones vital to human interests. Full restoration of the damage would not generally be practicable.

We have not followed the procedure usually applied to examining related questions, which has focused on the system's stability. This mathematical concept assumes that an attainable equilibrium is already present, and examines how the system reacts to subsequent departures from it, but clearly one should first look for factors which determine whether an equilibrium will be attained at all. (This is obviously true when local equilibrium about an equilibrium point is studied, but also applies to more complicated cases such as limit cycle behaviour, where a stable cycle loop cannot exist unless an attainable equilibrium point lies within it.) This is all the more advisable because, in a whole range of important cases, the existence of an attainable equilibrium effectively guarantees stability anyway⁴.

Here, then, we look for properties which distinguish a system with an attainable equilibrium. More specifically, we consider interacting variables $N_1, N_2, \dots$, varying with time, which can take only positive values. (Examples may include the 'populations' of animal, plant or chemical species, measured in numbers of individuals or (biomass).) Thus a set of mathematically possible equilibrium values $N_1^*, N_2^*, \dots$ can be attained

only if they are 'feasible'—that is, all $N_i^* > 0$ (ref. 5). Terms used here are 'species', 'populations' and 'feasible' equilibria.

This approach has a further substantial advantage: the model examined can be a very general one, so that any findings are widely applicable and far less dependent on the minutiae of a particular model specification. We write

$$dN_i/dt = f_i(N_i) [r_i + A_i(N_1, N_2, \dots, N_M)], \quad i = 1 \text{ to } M \quad (1)$$

Here the functions A_i describe the interactions between and within species; we will see that their form can be left almost completely open. The functions f_i are likewise left unspecified; they describe the species' reproductive (or decay) behaviour, as do the constants r_i . We require, however, that $f_i(0) = 0$, to exclude spontaneous creation.

In many types of system there are mechanisms by which the species can 'adapt' to each other, so that, over a long enough period of coexistence, the interactions take on a systematic character. However, because it is precisely the conditions for this initial coexistence which interest us, we exclude this possibility by requiring the species to interact randomly. We also exclude any species which, if left in isolation, would increase without limit, by requiring all species to be self-regulated.

The randomness of interaction will generally result in the disintegration of the assemblage as time goes on, with one or more species dying out. When M is large, only a small fraction of the solutions will survive intact, corresponding to rare configurations of the A_i . The nature of this survival can be mathematically various (local stability about N^* , limit cycle, strange attractor—see ref. 1, Ch. 2), and can generally be unravelled only if the A_i are fully known. However, for its survival to be possible, a system must first have a feasible equilibrium point. Thus, by studying the assemblages described by feasible solutions of equation (1), we are studying all possible candidates for survival as a persistent system.

Because we use numerical methods, definite probability distributions must be chosen. We first rewrite equation (1) formally as

$$dN_i/dt = f_i(N_i) \left[r_i + \sum_{j=1}^M A_{ij}^0 N_j \right] \quad (2)$$

Here the A_{ij}^0 are partial derivatives, and only mild restrictions need be imposed on the interactions A_i to justify rewriting in this form. (All first partial derivatives must be continuous; the first mean value theorem then gives equation (2), the derivatives being evaluated at a point which depends on the specific form of A_i but not on j .)

We now express the randomness of interaction by choosing

$$A_{ij}^0 = \pm g (i \neq j), \quad r_i = \pm 1 \quad (3)$$

Here g is a constant measuring interaction strength; all variates are independent and + or – with equal probability. For self-regulation, we use $A_{ii}^0 = -1$, all i .

In our computer studies, we started with a feasible solution for $M = 25$ variables, generated by the method of 'natural-selective elimination' (ref. 6). We then examined the structure of this solution, proceeding as follows: we selected q different species at random and eliminated them from the system; the resulting subsystem of $m = 25 - q$ species was then tested for feasibility. For a given value of m , 20 such subsystems were sampled (with replacement). This procedure was repeated for all subsystem orders from $m = 24$ to $m = 1$. In all, 100 systems of 25 species each were treated in this way, so that 2,000 subsystems of each order were examined. The value of g chosen was 0.16, corresponding to a May parameter⁷ $\gamma \equiv (M-1)^{1/2}g = 0.8$ in the 25-species system under study. The number of species with $r_i < 0$ was 2–11, with a mean of 5.33.

The results are summarized in Table 1. Line *a* shows the mean percentage of m -species subsystems found to be feasible, for $m = 1$ –24. (The s.e.m., as estimated from the sample variance, did not exceed 2%.) To appreciate how significantly large these percentages are, they should be compared with those in line *b*. These were derived from a control group of 100 initially

Table 1 Analysis of the feasibility structure of feasible and unfeasible solutions to equations (2) and (3)

Subsystem order (no. of species)	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
Mean % feasible:																								
<i>a</i> , In initially feasible system	51	31	21	15	11	9	8	8	6	7	7	7	9	9	9	12	15	19	22	29	39	46	61	80
<i>b</i> , In initially unfeasible system	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0.2	0.4	1	2	4	6	13	26	51

a Analyses the feasibility structure of 100 feasible solutions and *b* that of unfeasible solutions to equations (2) and (3), with scaled interaction strength $\gamma = 0.8$ ($g = 0.16$). A zero entry means that no feasible subsystem was found among the 2,000 tested.

unfeasible solutions ('non-survivors'), processed in exactly the same way. Even for $m = 16$, where the percentage feasible is least (6%), of the 2,000 subsystems examined, 122 were feasible, compared with the control group figure of zero.

To consider what these results mean, we should relate them to existing concepts of complex-system structure. Here, looking at the quantitative features of a definite model, we can confirm and significantly develop some of these mainly qualitative ideas. The concept is familiar, for example, of a persistent system built up from subsystem components which are themselves persistent. Thus the above results may be viewed as a concrete realization of, for example, Simon's 'Chinese box' picture⁸; but in further and important respects, it differs strikingly from all the previous realizations known to us.

The 'nesting' of subsystems is usually seen as achieved by linking them 'externally' to each other, so that k subsystems, each of order m , form a system of order km . Such an arrangement relates well to the needs of control theory, where an externally imposed negative feedback can stabilize each subsystem, and the links between subsystems can be made deliberately weak to preserve overall stability (see for example ref. 9, Ch. 1, 2). This concept seems particularly well adapted to subsystems which are 'mechanical', both in the everyday sense as well as Bertalanffy's more general one¹⁰.

However, the subsystems in the model above cannot be seen as just 'externally' linked in this way. On the contrary, the feasible subsystems of a given order overlap each other to an extraordinary degree. If we look once again at the least feasible case ($m = 16$), we easily calculate that each 25-species system contains, on average, over 120,000 feasible subsystems of order 16; a pair of these subsystems have, on average, over 10 species in common. Another striking comparison is provided by the feasible 5-species subsystems: the most economical Chinese box packing would need just five of these—the average above is over 15,000.

This overlapping suggests what might be termed a 'third level' on which to characterize a complex system. On the first level, we specify the single-species properties (intrinsic increase rate, and self-regulation parameter (s)) and on the second level, the pair interactions. The third level concerns the details of 'persistence structure', that is, the nesting and degree of overlap of component subsystems which are themselves viable. Such a characterization would be possible, and perhaps of some use, even when first- and second-level knowledge was deficient or altogether absent (as in catastrophe theory).

A further novelty in this model should be stressed—the strength of interaction involved. When the links between subsystems are external, stability is easiest to assure by keeping the interaction strength down to values within the 'weakly interacting' region. In our case, this corresponds to $\gamma < [M - 1]^{-1} = 0.04$. (The system is then biased towards stability by the Gersgorin criterion¹¹.)

The value used above is 20 times greater ($\gamma = 0.8$), therefore we are well into the 'significant' region. It is only here (Ashby's 'richly joined' case¹²) that system properties are primarily determined by the interactions; in the weaker regime, the behaviour of the isolated components is only weakly modified by the interactions.

These considerations suggest, then, that the nested, overlapping structure found in the model may be generally required if a truly autonomous ('natural') system is to persist in time, when

the interactions are numerous and strong enough to produce fully developed system behaviour. The robustness of such a structure, and the formidable difficulties in trying to reproduce it artificially (whether *ab initio* or after damage) when no control level is present, are apparent.

Received 23 April; accepted 7 October 1980.

1. May, R. M. *Stability and Complexity in Model Ecosystems* (Princeton University Press, 1973).
2. Odum, E. P. *Fundamentals of Ecology* (Saunders, Philadelphia, 1971).
3. Fox, S. W., & Dose, K. *Molecular Evolution and the Origin of Life* (Freeman, San Francisco, 1972).
4. Tregonning, K. thesis, Monash Univ. (1979).
5. Roberts, A. *Nature* **251**, 607 (1974).
6. Tregonning, K. & Roberts, A. *Nature* **281**, 563–564 (1979).
7. May, R. M. *Nature* **238**, 413–414 (1972).
8. Simon, H. A. in *Hierarchy Theory* (Braziller, New York, 1973).
9. Siljak, D. D. *Large-scale Dynamic Systems* (North-Holland, New York, 1978).
10. von Bertalanffy, L. *General System Theory* (Braziller, New York, 1969).
11. Marcus, M. & Minc, H. *A Survey of Matrix Theory and Matrix Inequalities* (Allyn & Bacon, Boston, 1964).
12. Ashby, W. R. *Design for a Brain* (Wiley, New York, 1960).

Why are fetal muscles slow?

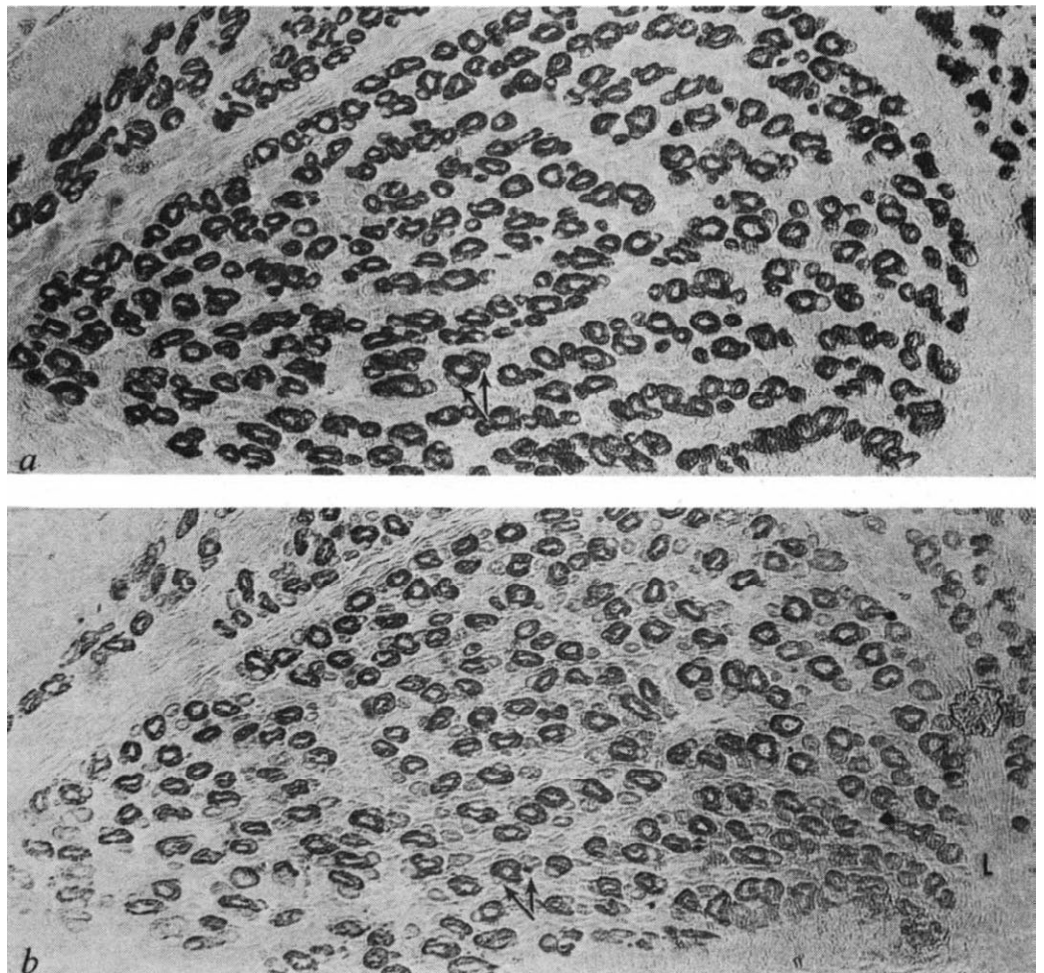
A. M. Kelly & N. A. Rubinstein

Departments of Pathobiology, School of Veterinary Medicine and of Anatomy, School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19174

Differentiating fast and slow mammalian muscles contract slowly at birth and increase their speed during the first few weeks of life^{1,2}. However, only small proportions of slow myosin light chains are found in early developing muscles and the fast type of light chains predominate^{3–7}. In addition, differentiating muscle contains unique, embryonic forms of myosin which may partially determine the early slow responses^{8–15}. The present study suggests additional reasons for these slow twitch times. Most skeletal muscles are initially formed from a small population of primary generation cells^{16,17} which are innervated by pioneering axons early in myogenesis¹⁸. Subsequently, numerous secondary generation cells develop along the walls of primary myotubes, then separate and become independent units of contraction. Using affinity-purified antibodies to fast and slow myosin⁵, it was found that most primary myotubes react with anti-slow myosin and are destined to become slow, Type I fibres. By contrast, secondary generation cells stain exclusively with anti-fast myosin and develop into Type II, fast fibres. We propose that primary myotubes constitute the fundamental motor units of the developing neuromuscular system and are responsible for early slow movements. Secondary generation cells become organized into large, fast motor units later in development, eclipsing the original slow response.

Serial, frozen sections were prepared from the extensor digitorum longus (EDL) and soleus muscle of fetal, neonatal and adult rats. At 16 days gestation primordia of the EDL and soleus stain with anti-fast myosin and the reaction with anti-slow is equivocal. Specification into fibre types in the EDL occurs between 17 and 18 days^{19,20}. This muscle is then composed of

Fig. 1 Serial sections of the EDL from an 18-day fetus stained by the indirect peroxidase anti-peroxidase method²⁰ with anti-fast myosin (*a*) and anti-slow myosin (*b*). The muscle primordium is mainly composed of large primary myotubes most of which stain with both anti-fast and anti-slow myosin. There are a few secondary generation cells attached to the walls of primary myotubes. All of these react strongly with anti-fast and weakly, if at all, with anti-slow myosin (arrows). The large myotubes containing anti-slow myosin are least numerous on the lateral side of the muscle (L). They are interpreted as components of fundamental motor units in the evolving neuromuscular system. $\times 210$.



150 to 200 large, primary myotubes most of which stain intensely with both anti-fast and anti-slow myosin (Fig. 1). There are numerous, less differentiated, secondary generation cells clustered around the walls of the primary myotubes, which stain intensely with anti-fast and weakly, if at all, with anti-slow myosin.

At 2 days post partum, comparable numbers of large fibres in the EDL stain with both anti-myosins (Fig. 2), but they are now widely separated by intervening fibres which stain only with anti-fast myosin. We interpret fibres staining with both anti-myosins as the primary generation cells, within which reactivity to anti-slow myosin is intense whereas the response with anti-fast is weak. This suggests that slow myosin is starting to dominate these cells and they are differentiating into Type I fibres (Fig. 2*b, c*). Studies of the adult EDL stained with anti-myosins and myofibrillar ATPase support this—there are approximately 200 Type I fibres amongst a total population of about 4,000 (ref. 21). Other fibres, which number at least 2,000 by 2 days, stain with anti-fast and not with anti-slow myosin. These correspond to the secondary generation cells which have now separated from primary cells and become independent

units of contraction. We interpret them as the precursors of Type II fibres. This pattern of assembly results in primary and secondary generation cells which are intermingled in a mosaic reminiscent of the checkerboard distribution of histochemically distinct fibres in adult muscle^{22,23}.

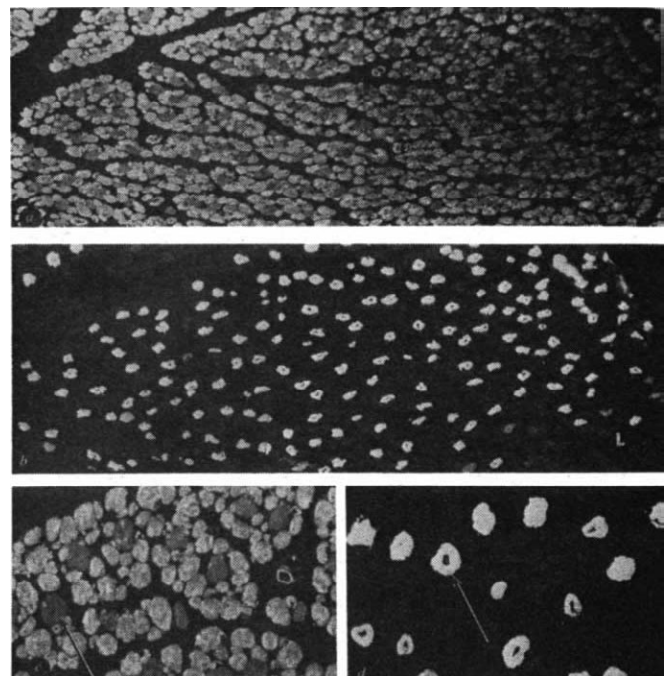


Fig. 2 Serial sections of the EDL at 2 days post partum stained by direct fluorescence with anti-fast myosin (*a, c*) and anti-slow myosin (*b, d*). Almost all fibres stain intensely with anti-fast myosin. There are a few, widely distributed fibres which stain weakly with anti-fast myosin (arrows, *c*). This selected population stains intensely with anti-slow myosin (arrows, *d*). These fibres are least numerous on the lateral side of the muscle (L) and correspond in distribution to the primary myotubes of the fetal EDL. They are interpreted as developing Type I fibres. Fibres staining only with anti-fast myosin are interpreted as secondary generation differentiating into Type II fibres. *a, b* $\times 75$; *c, d* $\times 140$.

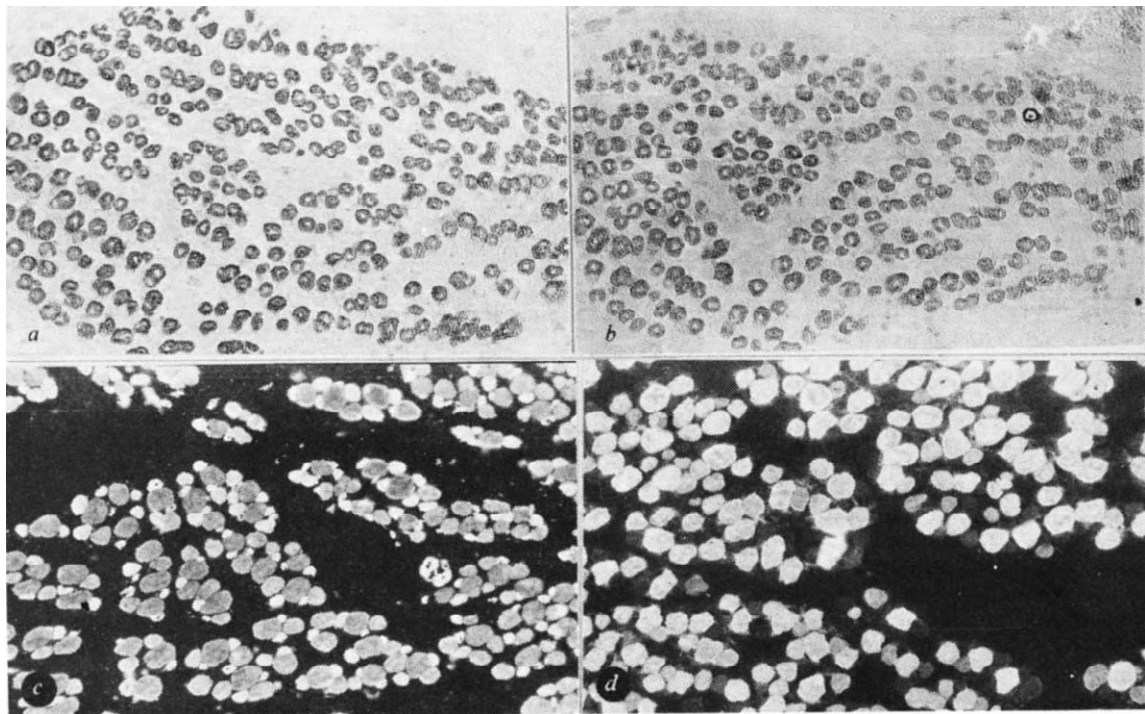


Fig. 3 Serial sections of the 18-day fetal soleus stained by the indirect peroxidase, anti-peroxidase method with anti-fast myosin (a) and anti-slow myosin (b). All fibres stain with both antibodies. There are very few secondary generation cells. At 2 days post partum (c, d, direct fluorescence) the intensity of staining with anti-fast myosin of large primary fibres has decreased but is intense with anti-slow myosin (d). In addition, there are now secondary generation cells developing along the walls of primary fibres all of which stain with anti-fast and not with anti-slow myosin. a, b $\times 170$, c, d $\times 190$.

Previous studies imply a close dependence of slow myosin synthesis on stimulation by the nervous system^{6,24,25}. In the 18-day EDL, primary myotubes are innervated and contain differentiating endplates²⁰. The appearance of slow myosin within these cells thus correlates with the development of neuromuscular relations. In addition, reflex hind limb movements develop in the fetus at approximately the same stage of gestation²⁶⁻²⁸, implying that motor unit activity is evolving. Hence we interpret the development of slow myosin within most primary myotubes as the reflection of a fundamental neuromuscular organization. In view of the muscle architecture, movements of the EDL at 18 days must be predominantly generated by primary myotubes containing significant proportions of slow myosin, which may be one reason why early contractions of the EDL are slow.

Secondary generation cells seem not to be separately innervated when in contact with primary myotubes^{6,29,30}. Many do not extend the entire length of the muscle^{16,31} but are attached to the primary cells by various forms of membrane junction. It is therefore doubtful that they can contract and relax independently. As the muscle matures, they gain independence^{16,31} and a separate nerve supply³². We propose that they are then added as developing fast motor units to the contractile machinery and as this occurs, they prevail over the early slow response and the muscle increases its speed of contraction.

Differentiation of the soleus follows the same pattern but the temporal sequence and ratio of primary to secondary fibres is different. Using our antibodies the 18-day fetal soleus is the only primordium amongst leg muscles in which almost all fibres stain uniformly with both anti-fast and anti-slow myosins (Fig. 3). In this muscle the results are consistent with those of Gauthier *et al.*³³. One reason for the uniformity of staining is that this primitive soleus is composed of myotubes which are small and have little variation in size. We interpret these as the primary myotubes of the soleus and there is little evidence of secondary orders of cells as in the EDL at this stage. Thus the soleus seems less developed than the EDL at 18 days.

Development of secondary generation fibres is evident in the soleus at 2 days post partum where their strong reactivity with anti-fast myosin contrasts with the less intense response of the larger primary generation fibres (Fig. 3c, d). As in the EDL, these secondary generation cells do not stain with anti-slow myosin whereas staining of primary fibres is intense. Slow myosin thus is the dominant species in the primary generation cells, which seem to be progressing towards Type I differentiation. Secondary generation cells are progressing towards Type II differentiation. They are as numerous as primary cells in the soleus at 2 days. Significantly, using myofibrillar ATPase stains, there are equal numbers of Types I and II fibres at 15 days post partum⁶.

These findings suggest that initial movements of the soleus, as in the EDL, result from contraction of primary fibres. They contain abundant slow myosin and slow rates of contraction may be anticipated. After birth, secondary generation, fast fibres appear and, once independent, these cells become separately innervated and are added as fast motor units to the original slow response. This may explain why the twitch time course of the soleus is initially slow and then becomes faster in the first few weeks of life. The work of Kugelberg³⁴ showing the transitions of whole motor units from fast to slow in the maturing animal accounts for the final slowing of the muscle.

This study was supported by NIH grants HL 15835 to the Pennsylvania Muscle Institute, NS 14332, BRSG 5 207, RR5464 and by a grant from the Muscular Dystrophy Association of America. We thank Dr Gerta Vrbova for first suggesting the interpretation of our fluorescence micrographs in terms of developing motor units.

Received 2 July; accepted 18 September 1980.

1. Buller, A. J., Eccles, J. & Eccles, R. J. *Physiol., Lond.* **150**, 399-416 (1960).
2. Close, R. J. *Physiol., Lond.* **173**, 74-95 (1964).
3. Sréter, F., Holtzer, S., Gergely, J. & Holtzer, H. *J. Cell Biol.* **55**, 586-594 (1972).
4. Pelloni-Mueller, G., Ermini, M. & Jenny, E. *FEBS Lett.* **67**, 68-74 (1976).
5. Rubinstein, N., Pepe, F. & Holtzer, H. *Proc. natn. Acad. Sci. U.S.A.* **74**, 4524-4527 (1977).
6. Rubinstein, N. & Kelly, A. *Dev. Biol.* **62**, 473-485 (1978).
7. Srogy, J. & Gutmann, E. *Pflügers Arch. ges. Physiol.* **369**, 85-89 (1977).
8. Trayer, I. P., Harris, C. I. & Perry, S. V. *Nature* **217**, 452-453 (1968).
9. Huszar, G. *Nature new Biol.* **240**, 260-264 (1972).

10. Sréter, F., Balint, M. & Gergely, J. *Devl Biol.* **46**, 317–325 (1975).
11. Whalen, R., Butler-Browne, G. & Gros, F. *J. molec. Biol.* **126**, 415–431 (1978).
12. Whalen, R., Butler-Browne, G., Sell, S. & Gros, F. *Biochemie* **61**, 625–632 (1979).
13. Whalen, R., Schwartz, K., Bouveret, P., Sell, S. & Gros, F. *Proc. natn. Acad. Sci. U.S.A.* **76**, 5197–5201 (1979).
14. Hoh, J. & Yeoh, A. P. S. *Nature* **280**, 321–323 (1979).
15. Rushbrook, J. & Stracher, A. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4331–4334 (1979).
16. Kelly, A. & Zacks, S. *J. Cell Biol.* **42**, 135–153 (1969).
17. Church, J. *J. Anat.* **105**, 419–438 (1969).
18. Bennett, M. & Pettigrew, A. *J. Physiol., Lond.* **241**, 515–545 (1974).
19. Rowlerson, A. *J. Physiol., Lond.* **301**, 19 (1980).
20. Rubinstein, N. & Kelly, A. *J. Cell Biol.* (in the press).
21. Carlson, B. & Gutmann, E. *Expl Neurol.* **58**, 82–93 (1976).
22. Ashmore, C., Robinson, D., Rattray, P. & Doerr, L. *Expl Neurol.* **37**, 241–255 (1972).
23. Kelly, A. & Schotland, D. in *Muscle Development and the Muscle Spindle* (ed. Banker) (Excerpta Medica, Amsterdam, 1972).
24. Salmons, S. & Sréter, F. *Nature* **263**, 30–34 (1976).
25. Salmons, S. & Vrbova, G. *J. Physiol., Lond.* **210**, 535–549 (1969).
26. East, E. *Anat. Rec.* **50**, 201–212 (1931).
27. Strauss, S. & Weddell, G. *J. Neurophysiol.* **3**, 358–369 (1940).
28. Windle, W., Minear, W., Austin, M. & Orr, D. *Physiol. Zool.* **8**, 156–185 (1935).
29. Kelly, A. & Zacks, S. *J. Cell Biol.* **42**, 154–169 (1969).
30. Kikuchi, T. & Ashmore, C. *Cell Tissue Res.* **171**, 233–251 (1977).
31. Ontell, M. *Anat. Rec.* **189**, 669–690 (1977).
32. Betz, W., Caldwell, J. & Ribchester, R. *J. Physiol., Lond.* **297**, 463–478 (1979).
33. Gauthier, G., Lowey, S. & Hobbs, A. *Nature* **274**, 125–129 (1978).
34. Kugelberg, E. *J. Neurol. Sci.* **27**, 269–289 (1976).

Anti-inflammatory steroids reduce tissue PG synthetase activity and enhance PG breakdown

P. K. Moore* & J. R. S. Hoult

Department of Pharmacology, King's College, Strand, London WC2R 2LS, UK

Anti-inflammatory steroids reduce prostaglandin (PG) synthesis in intact cells and isolated organs^{1–9} by interfering indirectly with the phospholipase(s) which release the polyunsaturated fatty acid precursors for both cyclooxygenase and lipoxygenase pathways. This action requires nucleic acid transcription and synthesis of new protein,^{10–12} and a soluble factor capable of inhibiting PG generation has been identified^{11,13}. However, it is not known whether these steroids affect either the actions or content of the enzymes of the PG system after administration *in vivo*, nor is it known if they affect PG-metabolizing enzymes. We show here that treatment of rats with anti-inflammatory steroids causes rapid changes in tissue activities of enzymes which synthesize and inactivate PGs, with apparent levels reduced and increased respectively.

Tissue activities of enzymes which synthesize and break down PGs were assayed by preparing homogenates of organs taken from animals pretreated with anti-inflammatory steroids or with vehicle alone. PG synthetase activity was generally measured by assaying the conversion of arachidonic acid to bioassayable PG-like material in freshly prepared 100,000g microsomal preparations^{14,15}, whereas that of the PG-metabolizing enzymes (here called 'prostaglandinases') was assayed by a radiochemical method using the 100,000g cytosolic supernatants and radiolabelled PGF_{2α} as substrate^{14,15}.

Treatment of rats with prednisolone, 0.5 and 40.0 mg per kg, for 10 days produced a dose-related decrease in activity of microsomal PG synthetase in caecum, kidney and lung; by contrast, the activities of prostaglandinase in these organs were increased (Fig. 1). An experiment investigating the time course of these reciprocal changes in the PG system showed that they occurred within 48 h, that is, after two subcutaneous injections of prednisolone, in caecum and kidney (Table 1) as well as in stomach (not shown).

Further experiments using prednisolone at 0.5, 5.0 and 40.0 mg per kg (given for 2 days only) also elicited dose-related decreases in apparent PG synthetase levels and increased prostaglandinase activity in stomach, caecum and kidney. The lowest

dose is similar to that recommended for a variety of clinical applications in man (10–100 mg per day)¹⁶. Several other anti-inflammatory steroids were tested at doses selected to reflect their established clinical potencies. As expected, these treatments caused a reduction in apparent microsomal PG synthetase activity, and in most cases also increased the prostaglandinase activity (Table 1). Correlations between anti-inflammatory potency and the anti-phospholipase actions of these steroids have been noted previously^{6,9}.

Taken together, the results suggest that these pharmacological effects of the glucocorticoids on the enzymes of the PG system might be related to their anti-inflammatory properties. To test this idea we treated rats with corticosterone, 17 α -methyltestosterone and tetrahydrocortisone (all at 20 mg per kg for 2 days), but they did not produce typical steroid-like reciprocal changes (Table 1). These therapeutically inactive compounds are closely related to hydrocortisone and may bind to glucocorticoid receptors, but are not agonists¹⁷. In most cases the treatments caused small changes in enzyme activities in the opposite direction (and represented in Table 1 as negative values), although in most cases the changes were not significant.

The simplest explanation for our findings is that the anti-inflammatory steroids modify cellular protein synthesis such that the amounts or levels of PG synthetase are decreased and those of PG-metabolizing enzymes increased. However,

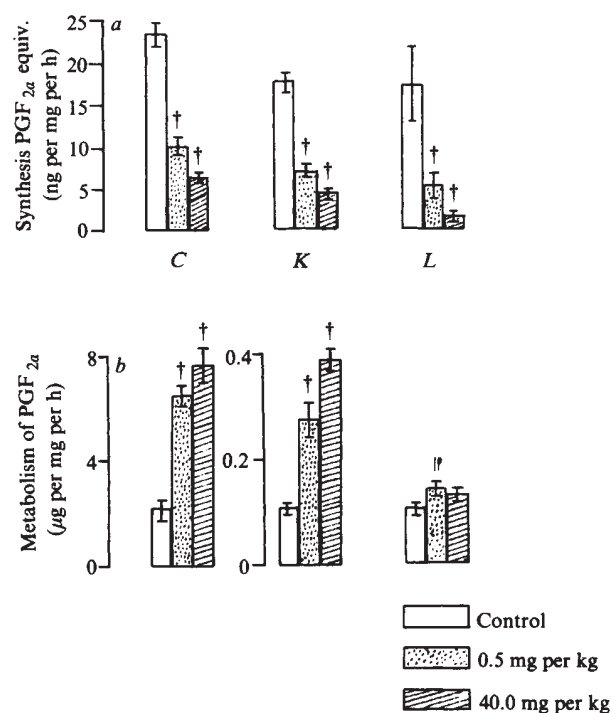


Fig. 1 Treatment of rats with prednisolone at 0.5 and 40.0 mg per kg for 10 days decreases activity (apparent level) in caecum (C), kidney (K) and lung (L) of microsomal PG synthetase (a), and increases activity of prostaglandin-metabolizing enzymes (b). Male Sprague-Dawley rats (five or six per group, 200 g) were injected with prednisolone sodium succinate or 0.2 ml vehicle (0.9% saline containing 0.5% CM-cellulose, 0.4% Tween 80 and 0.9% ethyl alcohol). Low-dose prednisolone animals appeared normal, but the growth and spleen size of rats treated with prednisolone 40 mg per kg were much reduced. However, histological examination of sections of stomach, lung and kidney did not reveal any differences between tissues from the three treatment groups. Organs were worked up individually as previously described^{14,15} and assayed for PG metabolism (100,000g supernatants incubated at 37 °C pH 7.5, with 5 mM NAD⁺, 0.05 μCi [9 β -³H]-PGF_{2α} and 10 μg ml⁻¹ PGF_{2α} and extracted for radio-TLC caecum 5 min, kidney 60 min and lung 90 min incubation), or PG synthesis (100,000g microsomes pooled and resuspended for 60 min incubation at 37 °C pH 7.5, with 10 μg ml⁻¹ arachidonic acid, 3 mM reduced glutathione and extracted for bioassay using PGF_{2α} as reference standard). Results show means $\pm$ s.e.m. for 6–12 determinations and are expressed in terms of protein content (not different between experimental groups for any single organ) using bovine serum albumin as standard. The significance of differences was evaluated using Student's *t*-test, and the key symbols are the same as those used in Table 1.

* Present address: Department of Pharmacology, Chelsea College, Manresa Road, London SW3 3TW, UK.

Table 1 Reciprocal changes in activity of enzymes of the prostaglandin system are induced *in vivo* within 2 days by anti-inflammatory steroids but not by their inactive analogues

Steroid	Dose (mg per kg)	Potency*	Duration of treatment (days)	% Reduction in tissue PG synthesis		% Enhancement of tissue prostaglandinase activity	
				Caecum	Kidney	Caecum	Kidney
Corticosterone	20	0.35	2	57.6 ± 6.8¶	5.0 ± 13.7 #	124.4 ± 33.3†	-14.7 ± 19.8 #
Triamcinolone	5	5	2	66.5 ± 5.4¶	49.8 ± 10.6§	95.3 ± 38.1¶	67.3 ± 4.2†
Dexamethasone	1	25	2	65.3 ± 4.3¶	54.1 ± 9.3†	226.2 ± 19.2†	-6.4 ± 24.8 #
Prednisolone	5	4	2	40.1 ± 4.6†	46.0 ± 5.2†	292.9 ± 20.6†	222.1 ± 42.1†
Hydrocortisone	20	1	2	58.6 ± 2.4†	57.5 ± 7.1†	228.4 ± 35.3†	185.7 ± 43.9†
Fludrocortisone	1.25	10	2	67.9 ± 1.3†	67.9 ± 2.3†	303.6 ± 55.8†	170.3 ± 39.3†
Tetrahydrocortisone	20	0	2	-64.9 ± 45.5 #	42.2 ± 3.5‡	-47.6 ± 4.7†	-22.7 ± 9.4 #
Cortexolone	20	0	2	-49.1 ± 7.9‡	23.2 ± 9.3 #	-13.7 ± 6.0 #	-19.8 ± 8.4 #
17 α -Methyltestosterone	20	?	2	-7.7 ± 20.3 #	-6.6 ± 9.4 #	-26.8 ± 11.0¶	-16.9 ± 8.6 #
Prednisolone	40	4	2	67.2 ± 4.2†	51.4 ± 5.8†	89.0 ± 7.2†	49.9 ± 5.6†
Prednisolone	40	4	4	66.6 ± 7.5†	45.5 ± 14.9¶	39.6 ± 21.1 #	52.4 ± 16.4¶
Prednisolone	40	4	8	62.9 ± 5.2†	41.2 ± 10.2§	32.9 ± 9.3§	88.5 ± 49.3 #

Groups of three or four rats were injected daily for 2–8 days with steroids as indicated. Doses refer to base: the steroids (all Sigma) were in the form corticosterone 21-acetate, triamcinolone acetate, dexamethasone acetate, prednisolone sodium succinate, hydrocortisone 21-acetate, fludrocortisone acetate, tetrahydrocortisone, cortexolone (=11-deoxycortisol) and 17 α -methyltestosterone. Enzyme activities are expressed as per cent change relative to vehicle-injected control rats. Results show mean $\pm$ s.e.m. for 4–10 determinations. Each block of results represents experiments with different batches of Sprague-Dawley rats.

* Relative anti-inflammatory potency, hydrocortisone = 1 (from ref. 16).

† $P < 0.001$; ‡ $P < 0.002$; § $P < 0.01$; ¶ $P < 0.02$; ¶ $P < 0.05$; # not significant relative to control treated rats, using Student's *t*-test.

alternative explanations are possible, and for this reason we refer to the changes in terms of 'activities' rather than 'levels'. It is recognized that many of the varied metabolic effects of glucocorticoids (and other steroid hormones) occur as a result of changes in the rates of synthesis of specific proteins^{18,19}.

It is unlikely that the PG synthetase results can be accounted for by a redirection of PG endoperoxide breakdown away from PGE₂ and PGF_{2 α} formation towards other non-spasmogenic prostanoid derivatives, because the reaction conditions (high substrate concentration, presence of reduced glutathione) favour the formation of the classical PGs (refs 20, 21). Furthermore, we obtained direct evidence from radiochemical experiments that the conversion of arachidonic acid to the various prostanoid products was reduced in freshly prepared homogenates of kidney, spleen and stomach prepared from rats treated for 2 days with prednisolone at 40 mg per kg. The overall percentage utilization of 0.5 μ g ml⁻¹ arachidonic acid (labelled with 0.4 μ Ci [1-¹⁴C]-arachidonic acid, Radiochemical Centre, specific activity 56.4 Ci mol⁻¹, no cofactor added) in 10-min incubations decreased from 41.0 $\pm$ 5.9 to 20.0 $\pm$ 4.9 in kidney, from 56.9 $\pm$ 5.3 to 10.9 $\pm$ 0.4 in spleen and from 26.7 $\pm$ 1.5 to 7.1 $\pm$ 2.6 in stomach, results obtained using six control and three steroid-treated animals.

We have shown that the activity of PG-metabolizing enzymes is increased in steroid-treated rats and believe that this may reflect increased tissue levels of prostaglandin 15-hydroxy-dehydrogenase (PGDH, the first enzyme in the PG degradative pathway²²). However, we cannot exclude the possibility that these changes can be partly explained by the presence in the supernatants of factor(s) which enhance PGDH activity (see accompanying paper²³) rather than by increases in PGDH levels. Nevertheless, this factor cannot be the steroid itself: in the range 1 μ M to 1 mM, prednisolone (four tests each at four concentrations) did not affect PGF_{2 α} breakdown in 60-min incubations of rat kidney 100,000g supernatant.

To confirm these findings, prostaglandinase activities in supernatants were also measured by bioassay of the rate of disappearance of 10 μ g ml⁻¹ PGE₂ and PGF_{2 α} from incubations of kidney and caecum. PGE₂ metabolism was more rapid than that of PGF_{2 α} , consistent with the known affinities of these PGs for PGDH²². A plot of disappearance of PGF_{2 α} activity against time in 100,000g supernatants of kidney prepared from rats pretreated for 2 days with prednisolone 40 mg per kg showed it to be accelerated (*t*_{1/2} ~ 10 min compared with ~ 32 min in control kidney). Both PGE₂ and PGF_{2 α} inactivation in caecum and kidney were enhanced in the steroid-treated rats.

Our results show that in the rat, pharmacological doses of anti-inflammatory steroids administered *in vivo* have a more pervasive 'anti-prostaglandin' effect than has previously been realised: they cause bidirectional enzymatic changes which may lead to reduced tissue PG synthesis and enhanced breakdown. We believe that the effects relate directly to their anti-inflammatory actions because PGs and related substances have predominantly pro-inflammatory properties^{24,25}; however, this effect is likely to be only one among many which contribute to the powerful anti-inflammatory spectrum of these compounds (see refs 16, 26). Although we do not understand how the changes in enzyme activities are brought about, it is probable that they reflect altered tissue levels of the enzymes concerned; this in turn suggests dependence on changes in cellular protein synthesis in the manner described for the steroid-induced release of anti-phospholipase factor^{11,13}.

We thank Dr J. E. Pike (Upjohn) for gifts of prostaglandins, the MRC and May & Baker Ltd for financial support and Dr J. G. Hoult and P. D. Brinck for assistance with histological studies.

Received 11 July; accepted 18 September 1980.

- Gryglewski, R. J., Panczenko, B., Korbut, R., Grodzinska, L. & Ocetkiewicz, A. *Prostaglandins* **10**, 343–355 (1975).
- Lewis, G. P. & Piper, P. J. *Nature* **254**, 308–311 (1975).
- Kantrowitz, F., Robinson, D. R., McGuire, M. B. & Levine, L. *Nature* **258**, 737–739 (1975).
- Tashjian, A. H., Voelkel, E. F., McDonough, J. & Levine, L. *Nature* **258**, 739–741 (1975).
- Floman, Y. & Zor, U. *Prostaglandins* **12**, 403–413 (1976).
- Nijkamp, F. P., Flower, R. J., Moncada, S. & Vane, J. R. *Nature* **263**, 479–482 (1976).
- Hong, S.-C. L. & Levine, L. *Proc. natn. Acad. Sci. U.S.A.* **73**, 1730–1734 (1976).
- Blackwell, G. J., Flower, R. J., Nijkamp, F. P. & Vane, J. R. *Br. J. Pharmac.* **62**, 79–89 (1978).
- Tam, S., Hong, S.-C. L. & Levine, L. *J. Pharmac. exp. Ther.* **203**, 162–168 (1977).
- Danon, A. & Assouline, G. *Nature* **273**, 552–554 (1978).
- Flower, R. J. & Blackwell, G. J. *Nature* **278**, 456–459 (1979).
- di Rosa, M. & Persico, P. *Br. J. Pharmac.* **66**, 161–163 (1979).
- Carnuccio, R., di Rosa, M. & Persico, P. *Br. J. Pharmac.* **68**, 14–16 (1980).
- Hoult, J. R. S. & Moore, P. K. *Br. J. Pharmac.* **68**, 719–730 (1980).
- Hoult, J. R. S. & Moore, P. K. *Clin. Sci.* **59**, 63–66 (1980).
- Haynes, R. C. Jr & Larner, J. in *The Pharmacological Basis of Therapeutics* 5th edn (eds Goodman, L. S. & Gilman, A.) 1472–1506 (Macmillan, New York, 1975).
- Samuels, H. H. & Tomkins, G. M. *J. molec. Biol.* **52**, 57–74 (1970).
- Baxter, J. D. & Forsham, P. A. *Am. J. Med.* **53**, 573–589 (1972).
- Thompson, E. B. & Lippmann, M. E. *Metabolism* **23**, 159–202 (1974).
- Chan, J. A., Nagasawa, M., Takeguchi, C. & Sih, C. J. *Biochemistry* **14**, 2987–2991 (1975).
- Cottee, F., Flower, R. J., Moncada, S., Salmon, J. A. & Vane, J. R. *Prostaglandins* **14**, 413–423 (1977).
- Hansen, H. S. *Prostaglandins* **12**, 647–679 (1976).
- Moore, P. K. & Hoult, J. R. S. *Nature* **288**, 271–273 (1980).
- Ferreira, S. H. & Vane, J. R. A. *Rev. Pharmac.* **14**, 57–73 (1974).
- Bonta, I. L. & Parnham, M. J. *Biochem. Pharmac.* **27**, 1611–1623 (1978).
- Parrillo, J. E. & Fauci, A. S. A. *Rev. Pharmac. Tox.* **19**, 179–201 (1979).

Pathophysiological states modify levels in rat plasma of factors which inhibit synthesis and enhance breakdown of PG

P. K. Moore* & J. R. S. Hoult

Department of Pharmacology, King's College, Strand, London WC2R 2LS, UK

We have shown recently that adaptive changes in the apparent amount of enzymes which synthesize and inactivate prostaglandins (PGs) occur in a reciprocal manner (see accompanying paper¹ and refs 2–4). For example, PG synthetase activity in several rat organs is reduced but that of PG-metabolizing enzymes ('prostaglandinases') is increased after treatment with anti-inflammatory steroids¹. In view of recent reports that the synthesis of PG-like substances may be influenced by plasma factors^{5–7}, we wondered whether our findings may be explained in whole or in part by the presence in varying amounts of substances which affect PG synthesis and inactivation in opposite directions. We show here that rat plasma contains a protein factor(s) which inhibits the synthesis of PGs and enhances their enzymatic breakdown *in vitro* and which we provisionally call prostaglandin 'reciprocal coupling factor' (RCF). Furthermore, RCF is rapidly released in response to anti-inflammatory steroids and its levels are altered in the two model pathophysiological states so far investigated.

The effects of plasma from male rats were tested in standardized assays of PG synthesis (in seminal vesicle microsomal preparations) and inactivation (using rat caecum 100,000g supernatants) as described in Fig. 1 legend. In both assays plasma was generally tested at final concentrations (v/v) of 0.5, 5.0 and 10.0%.

Low concentrations of rat plasma caused dose-related inhibition of PG synthesis in bovine seminal vesicle microsomes (Fig. 1a) and enhanced the enzymatic inactivation of PGF_{2α} in rat caecum supernatants (Fig. 1b). We suggest the provisional term prostaglandin 'reciprocal coupling factor' to designate the

plasma component(s) responsible for these effects. The dose-response curves obtained at the three concentrations tested were usually not linear (with characteristic upward inflections in 25 out of 34 synthesis-inhibition curves and in 29 of 30 metabolism-activation curves), suggesting that the observed effects might be caused by more than one plasma component.

Rat plasma at 5% (v/v) also inhibited the synthesis of bioassayable PG-like material in microsomal systems from four other species (see Fig. 1 legend). That the synthesis inhibition is not due to redirection of endoperoxide breakdown towards other products which cannot be detected by the bioassay (for example, the stable metabolites of prostacyclin or thromboxane A₂) was shown by incubating bovine seminal vesicle microsomes with radio-labelled arachidonic acid. In the presence of 0.5–10% plasma there was a progressive reduction in the utilization of arachidonic acid and in the amounts of each of the prostanoate products formed. The experiments did not distinguish whether inhibition of synthesis is due to depletion of substrate by binding (perhaps to albumin, as demonstrated previously^{8–10}), or to a specific inhibitory effect on the cyclooxygenase component of prostaglandin synthetase.

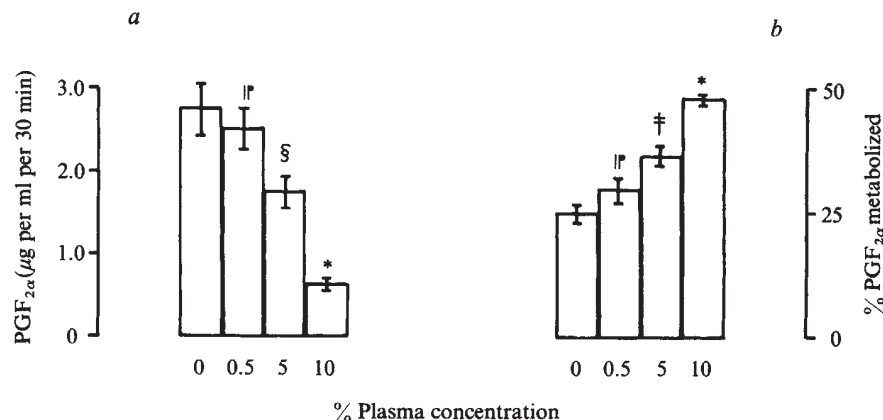
Enhancement of PGF_{2α} inactivation by rat plasma was shown in several other 100,000g supernatants using the radiochemical assay (Fig. 1 legend). To check these results, the time courses of breakdown of prostaglandins E₂ and F_{2α} in rat caecum supernatants were also measured by bioassay: as expected, PGE₂ was the better substrate (*t*_{1/2} ~ 1.6 min compared with ~ 5.6 min for PGF_{2α}), and the rate of disappearance of both PGs was enhanced in the presence of 5% plasma (*t*_{1/2} reduced to ~ 0.8 and ~ 2.9 min, respectively).

We wondered whether alterations in the amounts of RCF might contribute to the adaptive changes in apparent levels of PG-synthesizing and metabolizing enzymes which were observed previously in certain animal models of pathophysiological states^{1–4}.

Rats were injected subcutaneously (s.c.) for 2 days with prednisolone (5 mg per kg) or vehicle. As expected¹, apparent levels of PG synthetase and prostaglandinase in kidney, lung, stomach and caecum were reduced and increased, respectively (Fig. 2 legend). Blood samples were collected by cardiac puncture and the plasmas assayed for RCF activity. The dose-response curves in Fig. 2a and b show that the steroid-treated plasmas contained considerably larger amounts of synthesis-inhibiting and inactivation-enhancing activity than those from the control group. The results in Fig. 2a confirm those of Saeed

* Present address: Department of Pharmacology, Chelsea College, Manresa Road, London SW3 3TW, UK.

Fig. 1 Demonstration of prostaglandin 'reciprocal coupling factor' (RCF) in rat plasma by assays on PG synthesis (a) and breakdown (b). Synthesis assays: Bovine seminal vesicle microsomes (5 mg ml⁻¹ lyophilized powder, Miles) were incubated for 30 min at 37 °C in pH 8.0 100 mM Tris-HCl buffer with 10 µg ml⁻¹ arachidonic acid (Sigma) and 3 mM reduced glutathione to favour conversion to classical PGs, and the amounts of PG-like material assayed after extraction on the isolated rat fundus strip preparation in terms of PGF_{2α} equivalents. Bars show mean values ± s.e.m., *n* = 12–16. Breakdown experiments: Freshly prepared 100,000g supernatants of rat caecum homogenized in pH 7.4 50 mM phosphate buffer (final protein concentration 8.6 ± 0.2 mg ml⁻¹, *n* = 67) were incubated at 37 °C for 4–10 min as appropriate with 5 mM NAD⁺ and 10 µg ml⁻¹ PGF_{2α} containing 0.02–0.05 µCi [9β-³H]-PGF_{2α} (Radiochemical Centre, specific activity 15 Ci mmol⁻¹), and the extent of PGF_{2α} breakdown estimated by radiochromatography¹⁹. Results show mean values ± s.e.m., *n* = 4–8. RCF was found in both plasma and serum, was unaffected by the presence of heparin or if urethane or ether were given to anaesthetize the rats. Both synthesis-inhibiting and breakdown-enhancing properties exhibited similar characteristics: they were completely destroyed after 30 s boiling or pronase digestion (250 U ml⁻¹, 16 h at 4 °C), but unchanged by incubation for 5 min at 37 °C or 18 h at 4 °C, by dialysis or by storage at -20 °C and repeated thawing. Most activity was recovered after precipitation by ammonium sulphate and 5 min heating at 60 °C. In Figs 1 and 2 the significance of differences was evaluated by Student's *t*-test, and the following key is used: * *P* < 0.001, † *P* < 0.002, ‡ *P* < 0.01, § *P* < 0.02, || *P* < 0.05; ¶ not significant. With regard to others systems: 5% (v/v) rat plasma inhibits microsomal PG synthesis in frog skin by 60.8 ± 8.8% (*P* < 0.001), rat kidney 37.3 ± 6.8% (*P* < 0.01), rabbit kidney 59.7 ± 1.5% (*P* < 0.001) and guinea pig kidney 49.3 ± 2.1% (*P* < 0.001) (*n* = 10) and enhances PGF_{2α} inactivation in 100,000g supernatants of rat kidney by 54.4 ± 12.9% (*P* < 0.02), rat stomach 18.4 ± 2.2% (*P* < 0.01), rat lung 130.6 ± 22.5% (*P* < 0.01), guinea pig stomach 60.3 ± 15.8% (*P* < 0.001), guinea pig kidney 69.8 ± 11.9% (*P* < 0.01) and guinea pig colon 74.1 ± 6.5% (*P* < 0.001) (*n* = 5).



*et al.*⁵ showing that glucocorticoid treatment increases the ability of plasma to inhibit microsomal PG synthesis, but they did not examine organ enzyme levels or test the effects of plasma on PG degradation. A further experiment (not shown) revealed that RCF is released within 30 min of the intra-arterial injection of prednisolone (2.5 mg per kg) as determined by both synthesis-inhibition and metabolism-enhancement assays.

Furthermore, we confirmed (Fig. 2 legend) a previous report that the apparent PG synthetase content in kidneys of rats made diabetic with alloxan is decreased whereas prostaglandinase

activity is elevated³. Plasmas collected from alloxan-diabetic rats contained considerably more RCF than those from control animals (Fig. 2c, d).

We anticipated that it might be possible to induce adaptive changes in the opposite direction (increased PG synthesis and decreased breakdown) by treatment of rats with thyroxine. Earlier experiments showed that this reduces prostaglandinase activity in lung and kidney¹¹. Lungs, caeca, kidneys and blood samples were taken from rats injected s.c. for 2 days with 200 µg thyroxine. In all organs tested there was a decrease in prostaglandinase and increase in PG synthetase activity (Fig. 2 legend). The plasmas from the thyroxine-treated animals contained markedly less RCF activity than the control plasmas, according to both the microsomal synthetase and the caecum supernatant assays (Fig. 2e, f).

These data show that pathophysiological adaptations in apparent levels of enzymes of the PG system are indeed accompanied by 'appropriate' changes in the amounts of RCF, at least in the three examples studied. To what extent might the plasma factor be responsible for the changes in apparent enzyme activities? At present we cannot exclude the possibility that altered tissue prostaglandinase activities in the cell-free 100,000g supernatants may simply reflect the presence of different amounts of RCF taken up by the tissue or present in the blood entrapped in it, rather than differences in absolute enzyme levels. (Indeed, several authors¹²⁻¹⁵ have noted that cytosolic high-speed supernatants are inhibitory to microsomal PG synthesis, perhaps because they contain RCF). It is unlikely that the changes in apparent PG synthetase levels can be explained in terms of altered amounts of RCF because these assays are performed using resuspended microsomes separated from the cytosol, and in some pathophysiological models (such as diabetes, Fig. 2) changes were found in some organs but not in others. Additional experiments (not shown) verified that RCF does not significantly bind to microsomes, thus making it unlikely that the large differences in apparent levels of PG synthetase can be simply due to differences in amounts of adsorbed RCF.

These studies suggest that the activities of the enzymes of the PG system are under sensitive metabolic control and may be susceptible to regulation by rapidly releasable blood-borne factors. Saeed, Collier and co-workers^{5,15} have also identified a factor(s) capable of inhibiting microsomal PG synthesis *in vitro* (calling it 'endogenous inhibitor of prostaglandin synthesis'), and others have subsequently shown that serum or plasma may influence the amount or direction of arachidonate metabolism in both intact and broken cell preparations^{6,7,16-18}. However, ours is the first demonstration of alterations in plasma factors in model pathophysiological states and shows for the first time that the enzymatic destruction of PGs may also be influenced by blood-borne factors. These results reinforce our view that the functional control and adaptive versatility of the PG system are likely to depend on modulation of breakdown as well as regulation of synthesis.

Finally, it should be noted that the consequences of RCF action, if functionally effective *in vivo*, would be 'anti-homeostatic': thus, high levels would cause or enhance a PG deficiency whereas its absence would favour a PG excess. Either way, this may lead to pathological changes. It is therefore important to discover the identity and mechanism of action of RCF and to find out how and why it is released in the intact animal.

We thank the MRC, Pharmacia AB, May & Baker Ltd and the Central Research Fund of the University of London for financial support, and Drs J. E. Pike, R. L. Jones, H. O. J. Collier and K. Crowshaw for gifts of prostaglandins and seminal vesicles. E. Ramcharan provided technical help.

Received 11 July; accepted 22 September 1980.

1. Moore, P. K. & Houlst, J. R. S. *Nature* **288**, 269-270 (1980).
2. Houlst, J. R. S. & Moore, P. K. *Br. J. Pharmac.* **69**, 272-273 (1980).
3. Houlst, J. R. S. & Moore, P. K. *Clin. Sci.* **59**, 63-66 (1980).
4. Houlst, J. R. S., Moore, P. K., Marcus, A. J. & Watt, J. *Ag. Actions Suppl.* **4**, 232-244 (1979).
5. Saeed, S. A. *et al. Nature* **270**, 32-36 (1977).

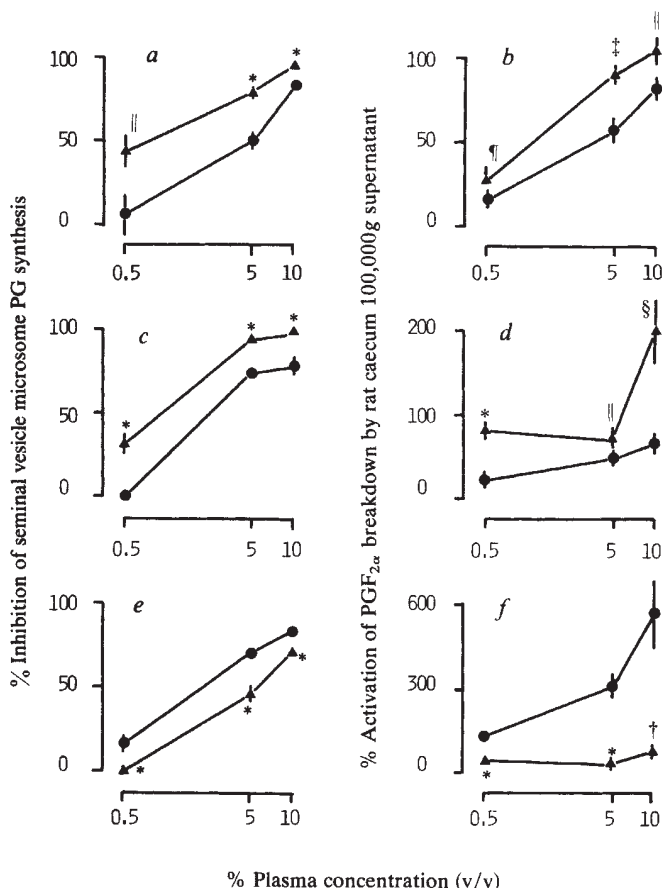


Fig. 2 Alterations in rat plasma RCF activity accompany the reciprocal changes in tissue enzyme activities of the PG system in three model pathophysiological states (prednisolone treatment, panels a and b; experimental diabetes, panels c and d; thyroxine administration, panels e and f). Panels a, c and e show inhibition of sheep seminal vesicle microsomal PG synthesis by RCF in rat plasma; panels b, d and f show activation of PGF_{2α} breakdown in rat caecum 100,000g supernatants by RCF in rat plasma. ● Denotes plasma from control and ▲ that from treated animals. Points show mean values $\pm$ s.e.m. (unless smaller than symbol) for 16-20 (synthesis) and 8-10 determinations (breakdown). See Fig. 1 legend for key to statistical evaluation and details of experimental methods. Reciprocal changes in tissue activities of enzymes which synthesize and inactivate PGs were seen after each of the three treatments as shown below. Prednisolone treatment (five Sprague-Dawley rats per group injected s.c. with prednisolone 5 mg per kg or vehicle for 2 days): PG synthetase activity in kidney, caecum, stomach and lung reduced by $45.0 \pm 5.0\%$ ($P < 0.02$), $38.0 \pm 7.6\%$ ($P < 0.05$), $29.5 \pm 12.2\%$ (not significant) and $46.0 \pm 5.9\%$ ($P < 0.05$) compared with activity in organs from vehicle-injected animals; prostaglandinase activity increased by $102 \pm 9.8\%$ ($P < 0.001$), $122.5 \pm 17.0\%$ ($P < 0.001$), $70.3 \pm 4.6\%$ ($P < 0.001$) and $57.8 \pm 12.9\%$ ($P < 0.01$) compared with control (see also ref. 1). Diabetic rats (five Wistar rats per group injected intravenously with alloxan 50 mg per kg or saline 3 days before death, blood sugar levels > 3.5 mg per ml in diabetic animals): kidney PG synthetase activity reduced by $48.9 \pm 5.7\%$ ($P < 0.002$) and prostaglandinase activity increased by $188.8 \pm 41.9\%$ ($P < 0.001$), lung activities not changed (see also ref. 3). Thyroxine administration (five Sprague-Dawley rats per group injected s.c. for 2 days with L-thyroxine 200 µg per kg): PG synthetase activity in kidney, caecum and lung enhanced by $169.5 \pm 13.2\%$ ($P < 0.001$), $30.4 \pm 7.1\%$ ($P < 0.05$), and $304.3 \pm 65.4\%$ ($P < 0.001$) compared with activity in control organs, prostaglandinase activity reduced by $48.3 \pm 5.1\%$ ($P < 0.01$), $29.9 \pm 2.6\%$ ($P < 0.01$) and $40.2 \pm 3.2\%$ ($P < 0.001$).

6. Herman, E. A., Yamamoto, M. & Rapoport, B. *J. cell. Physiol.* **100**, 401–406 (1979).
7. MacIntyre, D. E., Pearson, J. D. & Gordon, J. L. *Nature* **271**, 549–551 (1978).
8. Putnam, F. W. in *The Proteins* Vol. 3, 2nd edn (ed. Neurath, H.) 153–267 (Academic, London, 1965).
9. Bills, T. K., Smith, J. B. & Silver, M. J. *Biochim. biophys. Acta* **424**, 303–314 (1976).
10. Isakson, P. C., Raz, A., Denny, S. E., Wyche, A. & Needleman, P. *Prostaglandins* **14**, 853–871 (1977).
11. Moore, P. K. & Houlst, J. R. S. *Prostaglandins* **16**, 335–350 (1978).
12. Takeguchi, C., Kohno, E. & Sih, C. *Biochemistry* **10**, 2372–2376 (1971).
13. Rose, A. J. & Collins, A. J. *Prostaglandins* **8**, 271–283 (1974).
14. Wlodawer, P., Kindahl, H. & Hamberg, M. *Biochim. biophys. Acta* **431**, 603–614 (1976).
15. Collier, H. O. J., Denning-Kendall, P. A., McDonald-Gibson, W. J. & Saeed, S. A. in *Haemostasis, Prostaglandins and Renal Disease* (eds Remuzzi, G., Mecca, G. & de Gaetano, G.) (Raven, New York, in the press).
16. Hong, S. L. & Levine, L. *J. biol. Chem.* **251**, 5814–5816 (1976).
17. Remuzzi, G. *et al. Thromb. Res.* **13**, 1007–1015 (1978).
18. di Rosa, M., Capasso, F. & Mascolo, N. in *Advances in Inflammation Research* Vol. 1 (eds Weissmann, G., Samuelsson, B. & Paoletti, R.) 493–498 (Raven, New York, 1979).
19. Houlst, J. R. S. & Moore, P. K. *Br. J. Pharmac.* **61**, 615–626 (1977).

High-affinity anti-oestrogen binding site distinct from the oestrogen receptor

Robert L. Sutherland*, Leigh C. Murphy,
Ming San Foo, Michael D. Green
& Anne M. Whybourne

Ludwig Institute for Cancer Research, University of Sydney, Sydney,
New South Wales 2006, Australia

Zygmunt S. Krozowski

Department of Biochemistry, Royal Prince Alfred Hospital,
Camperdown, New South Wales 2050, Australia

Non-steroidal anti-oestrogens such as tamoxifen, CI 628, nafoxidine and clomiphene, are structurally related synthetic compounds that antagonize the effects of oestrogen on its target tissues^{1,2}, and this activity has led to the use of tamoxifen to treat advanced breast cancer³. All these compounds inhibit the binding of tritiated oestradiol to cytosol from oestrogen target tissues^{2,4–7}, suggesting that anti-oestrogens bind to the oestrogen receptor. This is supported by reports that in the rat uterus^{7–10} and dimethyl-benz (α)-anthracene (DMBA)-induced rat mammary carcinoma⁹, oestradiol and anti-oestrogens bind directly to the same number of saturable binding sites. Furthermore, oestrogens and anti-oestrogens are mutually competitive for binding to these sites^{8,10}. It has thus been generally accepted that the anti-oestrogens exert most of their effects through the specific oestrogen receptor. We now report a further high-affinity, anti-oestrogen binding site which may have a role in regulating the effects of non-steroidal anti-oestrogens.

Our evidence comes from comparison of the concentrations of high-affinity, saturable binding sites for oestradiol and anti-oestrogens in the same cytosol preparations, and from competition experiments with tritiated anti-oestrogens. Typical results of saturation analysis experiments with anti-oestrogens are shown in Fig. 1. The binding curves for the four oestrogen target-tissue cytosols were curvilinear (Fig. 1a) but could readily be resolved into a high-affinity, saturable binding site and non-specific binding^{11,12} (Fig. 1b). The straight-line Scatchard plot¹³ obtained by correction for non-specific binding (Fig. 1b) indicated a single high-affinity site or several saturable sites with similar affinities. No high-affinity sites were detected in the muscle cytosol (Fig. 1a).

Table 1 summarizes the apparent equilibrium dissociation constants (K_d) and concentrations (C) of high-affinity sites for oestradiol and tamoxifen in cytosol from eight oestrogen target tissues. With the exception of immature rat uterus, all these cytosols contained significantly more anti-oestrogen binding

sites than could be accounted for by oestrogen receptors alone. We have subsequently shown that the anti-oestrogen binding site is also present in immature rat uterus but was masked in the study reported here by a high concentration of oestrogen receptor¹⁴.

To test the ability of oestradiol to inhibit the binding of tritiated anti-oestrogen to its saturable binding sites, cytosol was

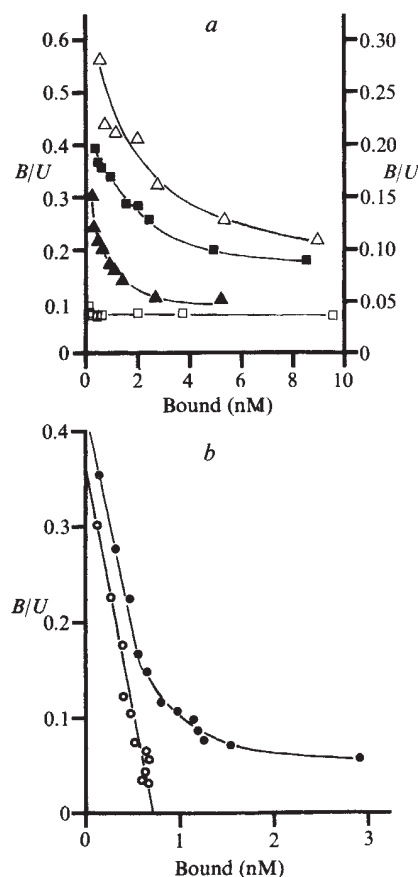


Fig. 1 Scatchard plots of anti-oestrogen binding to human endometrium, rat liver, chick oviduct, rat uterus and chick muscle cytosol. All tissues were homogenized with a Polytron and/or Teflon-glass tissue grinder, in 5–25 volumes (w/v) of 10 mM Tris-HCl buffer, pH 7.4, containing 1.5 mM EDTA and 0.5 mM dithiothreitol. A high-speed cytosol was prepared. Saturation analysis was performed at 4 °C by incubating 100 μ l of cytosol with a fixed concentration of tritiated tamoxifen (ICI, specific activity 19.5 Ci mmol⁻¹) or CI 628 (1.25 nM, Radiochemical Centre, Amersham, specific activity 8–25 Ci mmol⁻¹) and increasing concentrations of unlabelled ligand in the range of 1.25 nM–2 μ M. After 16 h of incubation, protein-bound and unbound ligand were separated by charcoal adsorption (0.5% charcoal for 30 min at 4 °C) and concentrations of bound and unbound ligand were calculated. The apparent equilibrium dissociation constant (K_d) and binding site concentration (C) of the saturable anti-oestrogen binding sites were calculated from measured bound and unbound ligand concentrations by a non-linear regression analysis technique¹¹ or by a graphical linearization technique¹². The former technique was preferred but in the only tissue (chick oviduct) where the techniques were compared on several samples both gave similar results. *a*, Binding of CI 628 to human endometrium cytosol (Δ) and rat liver cytosol ($\blacksquare$), and binding of tamoxifen to mature rat uterine cytosol ($\blacktriangle$) and chick skeletal muscle cytosol ($\square$). The scale of the left-hand ordinate applies to human endometrium only. *b*, Binding of CI 628 to chick oviduct cytosol before ($\bullet$) and after ($\circ$) correction for nonspecific binding. Tritiated CI 628 was greater than 97% pure at dispatch but was a 40:60 mixture of the *cis* and *trans* isomers. Tritiated tamoxifen was 88–90% pure on dispatch and was pure *trans* isomer. Purity was monitored continually by thin-layer chromatography and did not differ significantly from that at dispatch during the course of the experiments.

* To whom correspondence should be addressed.

Table 1 Binding parameters for interactions between oestradiol, tamoxifen and cytosol from eight oestrogen target tissues

Tissue	Mean cytosol protein conc. (mg ml ⁻¹)	No. of observations	Oestradiol		Tamoxifen		Tamoxifen/oestradiol	
			<i>K_d</i> (nM)	<i>C</i> (nM)	<i>K_d</i> (nM)	<i>C</i> (nM)	<i>K_d</i> (nM)	<i>C</i> (nM)
Rat uterus (immature)	1.85	10	0.17 ± 0.02	0.99 ± 0.07	2.54 ± 0.37	1.08 ± 0.08	14.9	1.09
Rat uterus (mature)	0.88	2	0.24 (0.23–0.25)	0.13 (0.12–0.14)	3.20 (2.28–4.11)	0.56 (0.55–0.56)	13.3	4.31
Rat liver	29.5	2	0.31 (0.16–0.46)	0.21 (0.14–0.28)	3.52 (2.97–4.06)	1.70 (1.66–1.74)	11.4	7.98
Rat kidney	18.9	2	0.14 (0.13–0.14)	0.12 (0.12–0.12)	5.60 (3.25–7.94)	0.80 (0.76–0.83)	40.0	6.67
Chick oviduct	3.4	6	0.07 ± 0.02	0.19 ± 0.01	9.82 ± 1.96	0.66 ± 0.24	140.3	3.47
Chick liver	20.3	2	0.93 (0.90–0.96)	0.11 (0.09–0.13)	3.62 (2.93–4.31)	0.50 (0.49–0.51)	3.9	4.55
Human endometrium	3.0	1	0.52	0.21	3.43	0.74	6.7	3.45
Human mammary carcinoma*	3.6	5	0.18 ± 0.07	0.17 ± 0.05	6.04 ± 1.60	1.47 ± 0.50	33.6	8.65

Experimental procedures are described in Fig. 1 legend. Binding-site concentration (*C*) is expressed as nmol of oestradiol or tamoxifen specifically bound per litre of reaction mixture, which was a twofold dilution of the cytosol. Data are presented as the mean ± s.e.m. or the mean plus range in parentheses.

* With oestrogen receptors.

incubated with tritiated CI 628 or tamoxifen and increasingly higher concentrations of oestradiol or anti-oestrogen. Figure 2 shows that the degree of inhibition by oestradiol varied with the tissue. With immature rat uterine cytosol, inhibition was complete^{7–10}, whereas in mature rat uterus and chick oviduct it was partial (Fig. 2). In human mammary carcinoma that had oestrogen receptors, oestradiol did not inhibit anti-oestrogen binding, whereas only non-specific binding occurred in chick muscle (Fig. 2). This can probably be explained by the relative affinities and concentrations of oestrogen and anti-oestrogen binding sites in cytosols, because those factors govern the amount of tritiated anti-oestrogen bound to the oestrogen receptor and thus the proportion of label displaced by oestradiol. For example, using data presented in Table 1 and elsewhere¹⁵, we calculated that only 5–7% of tritiated tamoxifen was bound to the oestrogen receptor in the mammary carcinoma cytosol, so that oestradiol could not displace significant amounts of label.

To assess the ligand specificity of the anti-oestrogen binding site we used cytosol from MCF 7 cells grown in 10% fetal calf serum because the relative concentration of saturable anti-oestrogen binding sites and oestrogen binding sites is greater than 50:1 (L.C.M. and R.L.S., unpublished data). As a result, essentially no tritiated anti-oestrogen was bound to the oestrogen receptor and competition curves represent the inhibition of binding of tritiated anti-oestrogen to anti-oestrogen binding site alone. Tamoxifen, CI 628 and nafoxidine all inhibited the binding of tritiated CI 628 to its saturable binding site (Fig. 3a), whereas natural and synthetic oestrogens, androgens, and progestins did not (Fig. 3a), indicating that the high-affinity anti-oestrogen binding site was not a sex steroid hormone receptor. To show that these results were not due to a contaminant of tritiated CI 628, similar experiments were performed with tritiated tamoxifen (Fig. 3b). Again anti-oestrogens (tamoxifen, CI 628, ICI 47,699, enclomiphene and *N*-desmethyltamoxifen) inhibited binding whereas the same steroids were ineffective.

The structural requirements for binding to the anti-oestrogen binding site are different from those for the binding of these anti-oestrogens to the oestrogen receptor. The *cis* isomers of tamoxifen and clomiphene (ICI 47,699 and zuclophene) have only 5–10% of the affinity of the *trans* isomers for the oestrogen receptor¹⁶ but both *cis* and *trans* tamoxifen have similar affinities for the anti-oestrogen binding site (Fig. 3b). Studies with 4-hydroxytamoxifen, a metabolite with 5–10 times the affinity of tamoxifen for the oestrogen receptor^{17,18}, showed that aromatic monohydroxylation had no effect on the affinity of tamoxifen for the anti-oestrogen binding site (A.M.W. and R.L.S., unpublished data). In contrast, demethylation in the

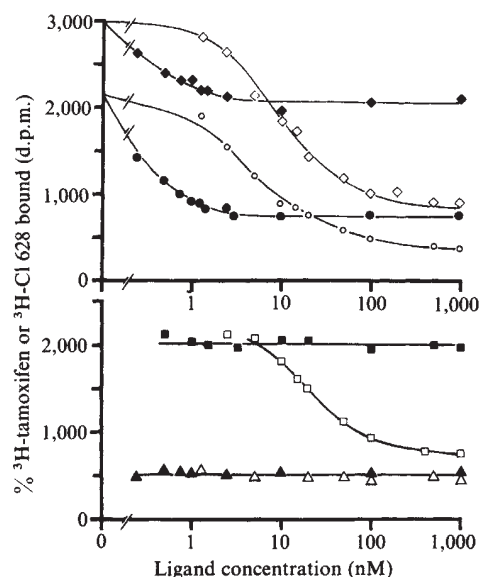


Fig. 2 Competition of oestradiol and anti-oestrogens (CI 628 or tamoxifen) for anti-oestrogen binding sites in chick oviduct, rat uterus, human mammary carcinoma with oestrogen receptors and chick muscle cytosol. Aliquots of cytosol were incubated for 16 h at 4 °C with 1.25 nM tritiated CI 628 or tamoxifen and increasingly high concentrations of unlabelled oestradiol or anti-oestrogen (tamoxifen or CI 628). Protein-bound radioactivity was separated by charcoal adsorption and the data were plotted as d.p.m. tritiated anti-oestrogen bound against log of the added ligand concentration. Results for chick oviduct (diamonds), mature rat uterus (circles), oestrogen receptor-positive human mammary carcinoma (squares) and chick skeletal muscle cytosols (triangles) are illustrated. Binding of tritiated anti-oestrogen in the presence of increasingly high concentrations of unlabelled oestradiol or anti-oestrogen are represented by solid and open symbols, respectively. Tamoxifen was the anti-oestrogen in the experiment with mammary carcinoma cytosol whereas in the other three experiments CI 628 was the anti-oestrogenic ligand. The protein concentrations of the cytosols are listed in Table 1 and the oestrogen receptor concentrations in fmol per mg protein were: chick oviduct, 112; mature rat uterus, 297; human mammary carcinoma, 119; and chick muscle, 0. For clarity the ordinate for the chick oviduct data has been expanded twofold, that is B_0 for chick oviduct = 1,500 d.p.m.

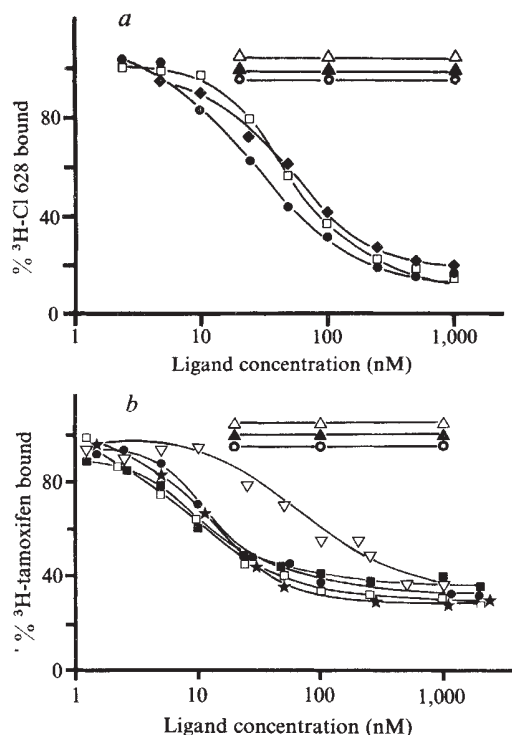


Fig. 3 Specificity of the saturable anti-oestrogen binding site in MCF 7 human mammary carcinoma cell cytosol. Tritiated CI 628 (a) or tamoxifen (b), cytosol and increasingly high concentrations of: tamoxifen (●), CI 628 (□), nafoxidine (◆), enclomiphene (★), ICI 47,699 (■), N-desmethyltamoxifen (▽); androgens: testosterone, 5α-dihydrotestosterone, R1881 (○); oestrogens: oestradiol, oestrone, diethylstilboestrol (▲), and progestins: progesterone, R5020 (△), were incubated for 16 h at 4°C and the protein-bound radioactivity was measured after adsorption on charcoal. Data are presented as the amount of tritiated anti-oestrogen bound in the presence of added ligand, expressed as a percentage of tritiated anti-oestrogen bound in the absence of added ligand against log of the concentration of added ligand.

alkylaminoethoxy side chain had little effect on receptor binding¹⁹ but markedly reduced the affinity for the anti-oestrogen binding site (compare tamoxifen with N-desmethyltamoxifen in Fig. 3b).

These data (Figs 1–3, Table 1), demonstrating an excess of high-affinity, saturable anti-oestrogen binding sites over oestrogen receptor sites and the inability of oestradiol to compete for all these sites, are evidence for an anti-oestrogen binding site. The site has a high affinity ($K_d = 3\text{--}10\text{ nM}$) for CI 628 and tamoxifen and is saturable at nanomolar concentrations of the drugs. It shows a surprising degree of specificity, binding a series of structurally related synthetic non-steroidal anti-oestrogens but not several steroid hormones (Fig. 3). There is also an indication of tissue specificity in that the site was present in all oestrogen target tissues studied but not in human mammary carcinoma that lacked oestrogen receptors¹⁵ nor in rat plasma and skeletal muscle. In addition we have shown that it is a protein, for binding activity can be destroyed by trypsin but not by lipase, RNase or DNase. The observation that its concentration changes during the oestrous cycle in rats²⁰ suggests that this intracellular binding protein is under hormonal control.

Our data are consistent with anti-oestrogen binding to the cytoplasmic oestrogen receptor as well as to the anti-oestrogen binding site. Indeed the competition curves illustrated in Fig. 2 and elsewhere⁷ demonstrate partial or complete inhibition of anti-oestrogen binding by oestradiol. We believe that anti-oestrogens bind to two saturable binding components *in vivo* which often cannot be resolved *in vitro* because of their relative

concentrations and affinities and the range of ligand concentrations used. However, experiments with human mammary carcinoma containing high titres of oestrogen receptor generated curves which could be resolved¹¹ into two saturable binding components (oestrogen receptor and anti-oestrogen binding site) in addition to non-specific binding²¹.

We have no evidence that this binding site mediates the effects of anti-oestrogen on target tissues. The possibility that it has a natural ligand and mediates the antagonistic effects of the drugs whereas agonist effects are mediated through the oestrogen receptor warrants investigation. In any case, the affinity and concentration of the anti-oestrogen binding site are such that it must bind a significant proportion of the drugs *in vivo* and so may regulate the amounts of anti-oestrogen available for binding to the oestrogen receptor.

We thank ICI, Warner Lambert/Parke-Davis, Merrell and Upjohn for anti-oestrogens, and Professor M. H. N. Tattersall for interest, encouragement and critical comment.

Received 14 March; accepted 25 September 1980.

- Katzenellenbogen, B. S. *et al. Rec. Prog. Horm. Res.* **35**, 259–300 (1979).
- Jordan, V. C., Dix, D. J., Naylor, K. E., Prestwich, G. & Rowsby, L. *J. Tox. envir. Hlth* **4**, 363–390 (1978).
- Mouridsen, H., Palshof, T., Patterson, J. & Battersby, L. *Cancer Treat. Rev.* **5**, 131–141 (1978).
- Horwitz, K. B. & McGuire, W. L. *J. biol. Chem.* **253**, 8185–8191 (1978).
- Lippman, M., Bolan, G. & Huff, K. *Cancer Res.* **36**, 4595–4601 (1976).
- Skidmore, J., Walpole, A. L. & Woodburn, J. *J. Endocr.* **52**, 289–298 (1972).
- Sutherland, R. L. & Foo, M. S. *Biochem. biophys. Res. Commun.* **91**, 183–191 (1979).
- Katzenellenbogen, B. S., Katzenellenbogen, J. A., Ferguson, E. R. & Krauthammer, N. J. *biol. Chem.* **253**, 697–707 (1978).
- Nicholson, R. I., Syne, J. S., Daniel, C. P. & Griffiths, K. *Eur. J. Cancer* **15**, 317–329 (1979).
- Capony, F. & Rochefort, H. *Mol. cell. Endocr.* **11**, 181–198 (1978).
- Sutherland, R. L. & Simpson-Morgan, M. W. *J. Endocr.* **65**, 319–332 (1975).
- Chamness, G. C. & McGuire, W. L. *Steroids* **26**, 538–542 (1976).
- Scatchard, G. *Ann. N.Y. Acad. Sci.* **51**, 660–672 (1949).
- Murphy, L. C. & Sutherland, R. L. *J. Endocr.* (in the press).
- Sutherland, R. L. & Murphy, L. C. *Eur. J. Cancer* (in the press).
- Sutherland, R. L. & Foo, M. S. in *Non-Steroidal Antioestrogens* (eds Sutherland, R. L. & Jordan, V. C.) (Academic, Sydney, in the press).
- Jordan, V. C., Collins, M. M., Rowsby, L. & Prestwich, G. *J. Endocr.* **75**, 305–316 (1977).
- Binart, N. *et al. Biochem. biophys. Res. Commun.* **91**, 812–818 (1979).
- Wakeling, A. E. & Slater, S. R. *Cancer Treat. Rep.* **63**, 1150 (1979).
- Faye, J. C., Lasserre, B. & Bayard, F. *Biochem. biophys. Res. Commun.* **93**, 1225–1231 (1980).
- Murphy, L. C. *et al. in Non-Steroidal Antioestrogens* (eds Sutherland, R. L. & Jordan, V. C.) (Academic, Sydney, in the press).

Stimulus–secretion coupling in rabbit neutrophils is not mediated by phosphatidylinositol breakdown

S. Cockcroft, J. P. Bennett & B. D. Gomperts

Department of Experimental Pathology,
School of Medicine, University College London, University Street,
London WC1E 6JJ, UK

In common with other cells which use intracellular Ca^{2+} to mediate specific cell function, when rabbit neutrophils are stimulated with specific agonists the rate of metabolism of phosphatidylinositol (PI) increases^{1–3}. This is normally measured as the incorporation of radioactive phosphate or inositol into PI, but these reactions are presumed to be secondary processes following the initial breakdown of pre-existing PI to diacylglycerol⁴. The radioactive labels are incorporated during the stepwise resynthesis of PI via phosphatidic acid (PA). It has been suggested that in the sequence of biochemical events, starting with the binding of the ligand to a receptor, and finally resulting in the expression of cellular activity, the breakdown of PI is an early event immediately directed by activation of the receptor^{1,4}. This could then control the increase in cytoplasmic Ca^{2+} and other processes dependent on this. Here we report an analysis of the temporal relationship between these phospholipid changes and cell stimulation. Our evidence suggests that in neutrophils, PI breakdown and PA labelling are both consequences and not causes of a rise in intracellular Ca^{2+} .

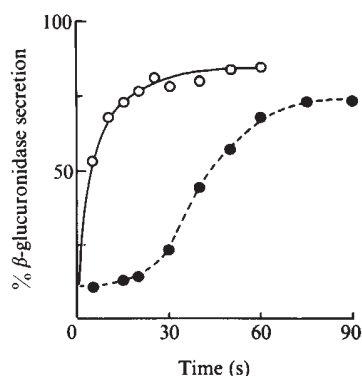


Fig. 1 Time course of β -glucuronidase secretion from rabbit neutrophils stimulated with 10^{-8} M fMet-Leu-Phe ($\circ$) and 5×10^{-6} M ionomycin ($\bullet$). Neutrophils were obtained from the rabbit peritoneal cavity 4–6 h after infusion of 250 ml 0.1% glycogen in saline³. They were suspended at 4×10^7 cells ml^{-1} in a buffered salt solution, pH 7.5, comprising 137 mM NaCl, 2.7 mM KCl, 1 mM MgCl_2 , 1.8 mM CaCl_2 , 20 mM HEPES, 5.6 mM glucose and 1 mg ml^{-1} bovine serum albumin. The cells were incubated at 37°C for 45 min. After addition of cytochalasin B ($5 \mu\text{g ml}^{-1}$), the cells were immediately added to fMet-Leu-Phe or ionomycin contained in small volumes of buffer. Samples containing $\sim 4 \times 10^7$ cells were quenched into ice-cold saline at the times indicated. The secreted β -glucuronidase released into the extracellular fluid was measured as previously described³. There was no detectable increase in the released β -glucuronidase in the absence of added ligands.

As agonists for neutrophil activation, we have used the synthetic tripeptide formylmethionyl-leucyl-phenylalanine (fMet-Leu-Phe) and ionomycin, which is a Ca^{2+} -carrying ionophore⁵. Activation of neutrophils normally results in an increase in cell motility, but in the presence of cytochalasin B ($5 \mu\text{g ml}^{-1}$), the cells secrete lysosomal enzymes (we measure β -glucuronidase) and the motile functions are suppressed⁶. Chemotaxis and secretion are known to be different expressions of the same cell-surface receptors⁷, and we have previously demonstrated that cytochalasin B does not significantly affect the binding of fMet-Leu-Phe to its receptor and the induced Ca^{2+} fluxes; nor does it alter the extent of PI labelling³.

Figure 1 shows the time course of β -glucuronidase secretion from cytochalasin B-treated neutrophils due to fMet-Leu-Phe (10^{-8} M) and ionomycin (5×10^{-6} M). The secretion due to the receptor-directed tripeptide is 90% complete within 10 s of adding the ligand, and so any early events which could dictate or modulate the course of the secretory process must be expressed within this time. Secretion due to the ionophore is characterized by a delay of about 20 s and is complete within about 60 s.

Figure 2 shows the changes in the radioactivity of PA and PI during 120 s following the addition of fMet-Leu-Phe to neutrophils which had been preincubated for 45 min in the presence of ^{32}P -phosphate to label phospholipids and the ATP pool. The radioactivity in these cells was quenched by direct addition to a mixture of chloroform/methanol (1:2), and the lipids were extracted and separated by TLC. Within 10 s of the addition of the ligand, the radioactivity of PI had declined by 30% whereas that of PA had increased slightly. Thereafter, the radioactivity in both lipids increased due to the incorporation of ^{32}P from ATP into the intermediates for PI resynthesis. In other experiments (data not shown), in which cells prelabelled with ^3H -glycerol were used, the radioactivity of PI declined to a similar extent within 10 s. In this case, the ensuing increase in radioactivity was not observed, because ^3H -glycerol can only enter PI as a consequence of *de novo* synthesis and, unlike ^{32}P -phosphate, is not incorporated during the resynthesis cycle¹.

Although the time dependence of PI breakdown certainly justifies its designation as an early event, our previous work showed that PI resynthesis (labelling) is unrelated to Ca^{2+} movements and the stimulation of secretion². Secretion due to fMet-Leu-Phe will occur in the absence of extracellular Ca^{2+} due to mobilization of Ca^{2+} from intracellular sources⁸, but in these circumstances we found that PI labelling does not occur². Table 1 shows that this is also true of PI breakdown. In addition, we show that when Ca^{2+} is introduced into the cell using ionomycin, secretion is accompanied by PI breakdown, and we have found that both responses are Ca^{2+} dependent. In this case, PI breakdown follows a slower time course similar to that of the ionomycin-stimulated secretion illustrated in Fig. 1 (results not shown).

We conclude that, in neutrophils, breakdown of PI is not essential for Ca^{2+} mobilization and secretion. Ionomycin and the receptor-directed agonist were equally effective in causing PI breakdown. Although the breakdown of PI is rapid, it is apparently a consequence and not a cause of the elevation of Ca^{2+} in the cytosol. Implicit in the results presented (Table 1) is the possibility that the concentration of intracellular Ca^{2+} required to induce PI breakdown might be higher than that needed for secretion. These results also exclude the recently suggested^{9,10} possibility that PA generated during resynthesis of PI could be the intrinsic ionophore which permits Ca^{2+} movement into agonist-stimulated cells. In neutrophils the appearance of PA is a late event and furthermore (unlike secretion), is dependent on extracellular Ca^{2+} (see Table 1).

There is a close correlation between cell activation processes mediated by an elevation of cytosol Ca^{2+} and the stimulation of PI metabolism^{1,4,11}. The possibility that PI breakdown could precede and thus regulate the fluxes of Ca^{2+} which lead to

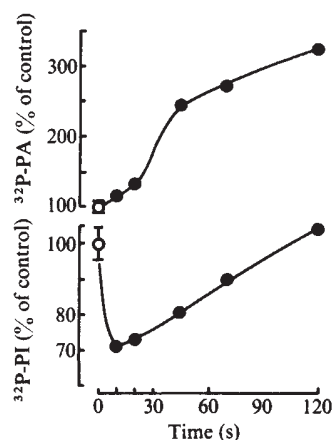


Fig. 2 Time course of the changes in radioactivity of PI and PA after addition of fMet-Leu-Phe to neutrophils. Neutrophils (4×10^7 cells ml^{-1}) were incubated at 37°C for 45 min with $50 \mu\text{Ci ml}^{-1}$ ^{32}P -Pi. Four samples (1 ml) were then quenched directly into 3.75 ml of chloroform/methanol (1:2) (zero time control) and the remainder of the cells were immediately added to fMet-Leu-Phe (final concentration 10^{-8} M) contained in a small volume. Samples were taken and quenched at the times indicated. The lipids were extracted¹³ and separated by TLC as previously described¹⁴. The lipids were identified by staining with iodine vapour and by autoradiography. The spots containing PI and PA were scraped into vials and the radioactivity measured by liquid scintillation counting. Open symbols indicate zero time control, mean $\pm$ s.e.m., $n = 4$; PA = $1,890 \pm 110$ c.p.m.; PI = $2,470 \pm 100$ c.p.m. There was no detectable increase in the radioactivity of these two lipids in control samples during the 2-min period of incubation. A similar kinetic pattern of phospholipid changes was obtained when cytochalasin B-treated neutrophils were used, and the cells quenched with ice-cold saline. In 12 independent experiments, using fMet-Leu-Phe at 10^{-8} M, we found that at 10 s breakdown of labelled PI was $26 \pm 2\%$ and the increment in PA labelling was $32 \pm 5\%$.

Table 1 Effects of Ca^{2+} on agonist-induced secretion and changes in radioactivity of ^{32}P -labelled phosphatidylinositol and phosphatidic acid

	c.p.m. in PI (% of control)	c.p.m. in PA (% of control)	β -glucuronidase secretion (% of total)
a Control	100 $\pm$ 1	100 $\pm$ 1	3 $\pm$ 1
fMet-Leu-Phe	107 $\pm$ 3	102 $\pm$ 7	15 $\pm$ 1
Control + Ca^{2+}	100 $\pm$ 1	100 $\pm$ 3	4 $\pm$ 1
fMet-Leu-Phe + Ca^{2+}	69 $\pm$ 5	153 $\pm$ 4	60 $\pm$ 1
b Control + Ca^{2+}	100 $\pm$ 2	100 $\pm$ 3	4 $\pm$ 1
Ionomycin + Ca^{2+}	78 $\pm$ 2	137 $\pm$ 3	50 $\pm$ 3

a, Effect of Ca^{2+} on secretion and changes in radioactivity of PI and PA after 10-s exposure of cytochalasin B-treated neutrophils to fMet-Leu-Phe. b, Secretion and changes in the radioactivity of PI and PA after 45-s exposure of cytochalasin B-treated neutrophils to ionomycin + Ca^{2+} . Neutrophils (4×10^7 cell ml^{-1}) were incubated for 45 min in a Ca^{2+} -free buffer (EGTA 10 μM) containing ^{32}P . After the addition of cytochalasin B (5 $\mu\text{g ml}^{-1}$), three pairs of samples (1 ml) were transferred in succession to tubes containing agonist (fMet-Leu-Phe 10^{-8} M or ionomycin 5×10^{-6} M) and buffer, with or without Ca^{2+} (1.8 mM); these were quenched after 10 s (fMet-Leu-Phe) or 45 s (ionomycin) by addition of 5 ml of ice-cold saline. The whole operation was complete within 2 min. After centrifugation at 4 $^{\circ}\text{C}$, the supernatant was sampled for estimation of β -glucuronidase and the cell lipids were extracted and analysed as described in Fig. 2 legend. Ca^{2+} had no effect on secretion and phospholipid changes in the absence of fMet-Leu-Phe. Results are expressed as mean $\pm$ s.e.m., $n = 3$. Concentration of Ca^{2+} in Ca^{2+} -free buffer (measured with a Ca^{2+} electrode) was less than 5 μM .

activation has been a popular concept because in many of the earlier systems studied, PI responses seemed to be independent of extracellular Ca^{2+} (ref. 2). If breakdown of PI has a universal role in cell activation processes, our results with neutrophils, and those of others with platelets¹², indicate that this is not in the regulation of Ca^{2+} fluxes.

This investigation was supported by grants from the Wellcome Trust and the MRC.

Received 21 April; accepted 30 September 1980.

1. Michell, R. H. *Biochim. biophys. Acta* **415**, 81–147 (1975).
2. Cockcroft, S., Bennett, J. P. & Gomperts, B. D. *FEBS Lett.* **110**, 115–118 (1980).
3. Bennett, J. P., Cockcroft, S. & Gomperts, B. D. *Biochim. biophys. Acta* **601**, 584–591 (1980).
4. Michell, R. H., Jafferji, S. S. & Jones, L. M. *Adv. exp. Med. Biol.* **83**, 447–465 (1977).
5. Bennett, J. P., Cockcroft, S. & Gomperts, B. D. *Nature* **282**, 851–853 (1979).
6. O'Flaherty, J. P. & Ward, P. A. *Semin. Hematol.* **16**, 163–174 (1979).
7. Showell, H. J. *et al. J. exp. Med.* **143**, 1154–1169 (1976).
8. Naccache, P. H., Volpi, M., Showell, H. J., Becker, E. L. & Sha'afi, R. I. *Science* **203**, 461–463 (1979).
9. Salmon, D. M. & Honeyman, T. W. *Nature* **284**, 344–345 (1980).
10. Putney, J. W., Weiss, S. J., van de Walle, C. M. & Haddas, R. A. *Nature* **284**, 345–347 (1980).
11. Gomperts, B. D., Cockcroft, S., Bennett, J. P. & Fewtrell, C. M. S. *J. Physiol., Paris* **76**, 383–393 (1980).
12. Bell, R. L. & Majerus, P. W. *J. biol. Chem.* **255**, 1790–1792 (1980).
13. Michell, R. H. & Jones, L. M. *Biochem. J.* **138**, 47–52 (1974).
14. Skipski, V. P., Peterson, R. F. & Barclay, M. *Biochem. J.* **90**, 374–378 (1964).

The possible functional significance of phosphatidylethanolamine methylation

Dennis E. Vance & Ben de Kruijff*

Department of Biochemistry, University of British Columbia, Vancouver, British Columbia, Canada V6T 1W5

The conversion of phosphatidylethanolamine to phosphatidylcholine was first reported in rat liver microsomes by Bremer and Greenberg¹. The reaction requires three successive methylations of the ethanolamine moiety by *S*-adenosylmethionine. The highest activity for this enzyme has been found in liver microsomes and the enzyme from rat liver was recently solubilized and partially purified². Although the activity is highest in liver, it is the minor pathway in this tissue for the synthesis of phosphatidylcholine³. The enzyme has recently been identified in bovine adrenal medulla⁴ and reports exist on the activity of phosphatidylethanolamine methyltransferase in erythrocyte ghosts⁵, reticulocyte ghosts^{6,7} and mammary gland membranes⁸. In view of the relatively minor importance of the

methylation pathway in phosphatidylcholine biosynthesis in liver, we were intrigued by the reports of significant physiological changes attributed to phospholipid methylation^{6–9}. Calculations reported here show that this enzymatic activity in reticulocytes, erythrocytes and mammary gland membranes is 0.1% of that observed in liver microsomes. Furthermore, in conditions where marked changes in microviscosity of the erythrocyte membrane were observed, only extremely small amounts of phosphatidylethanolamine were methylated⁹. For these and other reasons, there is considerable doubt that methylation of phosphatidylethanolamine could account for the many physiological responses attributed to this activity.

Table 1 compares the activity for *N*-methylation of phosphatidylethanolamine in rat liver microsomes with the highest activity reported for several other membrane preparations. The activity in the microsomes from adrenal medulla is 100-fold lower than in liver microsomes and 1,000-fold lower in the three other membrane preparations. In the presence of exogenously added phosphatidyl-*N*-monomethylethanolamine, the activity in liver microsomes and in erythrocyte ghosts was 0.58 (ref. 2) and 0.0003 nmol (Fig. 1 of ref. 5) ^3H -methyl transferred per min per mg protein, respectively. In view of the very low enzymatic activities of the last three membrane preparations listed in Table 1, we question whether or not the activity is associated with these membranes. Perhaps the activity is due to cellular impurities in the preparation.

Because such low methylation activity was observed with the erythrocyte membrane, we decided to calculate the number of molecules of phosphatidylethanolamine methylated during these incubations. In one reported experiment, erythrocyte ghosts were incubated for 1 h at 37 $^{\circ}\text{C}$ with 100 μM *S*-adenosylmethionine⁹. During this incubation 4.4 pmol of phosphatidylethanolamine were methylated per mg of membrane protein. The ratio of protein to lipid in rat erythrocytes is 1.1 (ref. 10). Thus, there would be 0.91 mg lipid per mg protein, and on a molar ratio half of this lipid (0.62 mg) is polar lipid¹⁰. Most, if not all, of the polar lipid is phospholipid. The phospholipids of the rat erythrocyte include 47.5% phosphatidylcholine and 21.5% phosphatidylethanolamine¹¹. From these data and the assumption of 750 and 800 as the respective molecular weights of phosphatidylethanolamine and phosphatidylcholine, we can calculate that rat erythrocyte membranes contain 173,000 pmol of phosphatidylethanolamine and 362,000 pmol of phosphatidylcholine per mg protein. During the incubation described above, 2.0 pmol of phosphatidyl-*N*-monomethylethanolamine, 1.2 pmol of phosphatidyl-*N,N*-dimethylethanolamine and 1.2 pmol of phosphatidylcholine were apparently formed⁹. Thus, 0.0012% of the phosphatidylethanolamine was converted to the monomethyl derivative. Of the phosphatidylcholine molecules present, only 0.00033% were formed via methylation of phosphatidylethanolamine in the 1-h incubation. Despite these extremely small amounts of lipids converted or formed, a relatively large change in microviscosity from 1.62 to 1.09 P was observed in the same experiment. Therefore, it seems highly unlikely that the changes in membrane microviscosity could have been caused by methylation of phosphatidylethanolamine. Some other explanation must be sought for the observed changes in microviscosity. This conclusion is supported by the studies of Schroeder *et al.*¹² with LM cells. There was no change in the fluidity of the membranes or membrane lipid (as measured by several different probes) when the polar head group of the phospholipid in the cell or subcellular membranes was altered by more than 50%.

Another area of concern in the reports on methylation of phosphatidylethanolamine is the extremely small amounts of ^3H incorporated into the lipids. There is a serious possibility that the ^3H radioactivity originates from contaminating compounds in the *S*-adenosyl-L-[*Me*- ^3H]methionine. For instance, in the

* Permanent address: Department of Molecular Biology, Padualaan 8, de Uithof, Utrecht, The Netherlands.

Table 1 Comparison of various enzymatic activities for methylation of phosphatidylethanolamine

Source of enzyme	Enzyme activity (nmol ³ H-methyl transferred per min per mg protein)
Rat liver microsomes (ref. 2)	0.45
Bovine adrenal medulla microsomes (Table 1 of ref. 4)	0.0040
Rat erythrocyte ghosts (Table 2 of ref. 5)	0.0005
Rat reticulocyte ghosts (Fig. 3 of ref. 6)	0.0001
Mammary gland membranes (Table 3 of ref. 8)	0.0003

Information in parentheses indicates the source of the data.

experiment reported in Table 1 of ref. 5, only 0.05% of the radioactivity is incorporated into phosphatidyl-*N*-monomethyl-ethanolamine. The radiopurity of the *S*-adenosyl-*L*-[*Me*-³H]methionine as supplied by New England Nuclear is approximately 99% by the time this compound is used. Thus, the amount of contaminating ³H-labelled compounds is many times higher than the actual amount of ³H incorporated. Therefore, it is essential that for each system quantitative data are provided which show that the only radioactivity incorporated in the membrane is that of ³H-methyl in the polar head group of the methylated phospholipid. In the reports^{6-8,13-17} that suggest a link between physiological effects and phospholipid methylation, such studies are lacking. In addition, these experiments are subject to the same criticisms as detailed above. Hence, we are not convinced that methylation of phosphatidylethanolamine is related to any physiological response or change in membrane fluidity.

We thank the MRC, the Canadian Heart Foundation and the BC Health Care Research Foundation for the support of B.K. while in Canada. This work was supported by grants from the MRC, BC Heart Foundation and NIH.

Received 18 March; accepted 18 September 1980.

- Bremer, J. & Greenberg, D. M. *Biochim. biophys. Acta* **46**, 205-216 (1961).
- Schneider, W. J. & Vance, D. E. *J. biol. Chem.* **254**, 3886-3891 (1979).
- Sundler, R. & Åkesson, B. *J. biol. Chem.* **250**, 3359-3367 (1975).
- Hirata, F., Viveros, O. H., Diliberto, E. J. & Axelrod, J. *Proc. natn. Acad. Sci. U.S.A.* **75**, 1718-1721 (1978).
- Hirata, F. & Axelrod, J. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2348-2352 (1978).
- Strittmatter, W. J., Hirata, F. & Axelrod, J. *Science* **204**, 1205-1207 (1979).
- Hirata, F., Strittmatter, W. J. & Axelrod, J. *Proc. natn. Acad. Sci. U.S.A.* **76**, 368-372 (1979).
- Bhattacharya, A. & Vonderhaan, B. K. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4489-4492 (1979).
- Hirata, F. & Axelrod, J. *Nature* **275**, 219-220 (1978).
- Quinn, P. J. *The Molecular Biology of Cell Membranes*, 31 (Macmillan, London, 1976).
- van Deenen, L. L. M. & de Gier, J. in *The Red Blood Cell* Vol. 1 (ed. Surgenar, D. M.) 147-211 (Academic, London, 1974).
- Schroeder, F., Holland, J. F. & Vagelos, P. R. *J. biol. Chem.* **251**, 6747-6756 (1976).
- Pike, M. C., Kredich, W. M. & Snyderman, R. *Proc. natn. Acad. Sci. U.S.A.* **76**, 2922-2926 (1979).
- Hirata, F., Axelrod, J. & Crews, F. T. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4813-4816 (1979).
- Mato, J. M. & Marin-Cao, D. *Proc. natn. Acad. Sci. U.S.A.* **76**, 6106-6109 (1979).
- Strittmatter, W. J. *et al. Nature* **282**, 857-859 (1979).
- Hirata, F., Corcoran, B. A., Venkatasubramanian, K., Schiffman, E. & Axelrod, J. *Proc. natn. Acad. Sci. U.S.A.* **76**, 2640-2643 (1979).

Reply from J. Axelrod and F. Hirata

Section on Pharmacology, Laboratory of Clinical Science,
National Institute of Mental Health, Bethesda, Maryland 20205

From papers published in our own and other laboratories, Vance and de Kruijff make calculations showing that phospholipid methylation activities observed in a variety of tissues and cells are very low. They then argue that because of these low enzyme activities, all the reported findings on the association between phospholipid methylation and biological responses are due to an impurity in *S*-adenosyl-*Me*-[*Me*-³H]methionine.

Vance and de Kruijff have not provided any experimental evidence for this assumption.

The reports on the methylation of membrane lipids satisfy the biochemical criteria of enzymatic reactions¹⁻⁴: (1) no significant activity in a heated preparation, (2) saturated substrate concentration-velocity curves, (3) pH dependency, (4) subcellular localization, (5) sensitivity to proteolytic enzymes, (6) stimulation by exogenous phospholipid substrates and cofactors and (7) competitive inhibition by *S*-adenosylhomocysteine. All the methylated lipids arising from *S*-adenosyl-[*Me*-³H]methionine were identified by TLC with several solvent systems, chemical and enzymatic degradation and column chromatography with ion-exchange resin of products from chemical degradation^{1,2,4-6}. These observations exclude the possibility of measuring the impurities as phospholipid methylation. The alternative explanation is that an impurity of ³H-*S*-adenosylmethionine has the capacity to transmethyate an acceptor lipid and an exceedingly high affinity (K_m in 10^{-9} M range) for the reaction. In experiments using intact cells (mast cells^{5,7}, basophils⁸, neutrophils⁹, astrocytoma cells¹⁰ and lymphocytes¹¹), [*Me*-³H]methionine, a precursor of *S*-adenosylmethionine, was used for the assay of phospholipid methylation. Therefore, the impurity should also be formed in intact cells from methionine and the transfer of the methyl group should be blocked by methyltransferase inhibitors such as *S*-adenosylhomocysteine and its analogues⁷⁻¹¹. It would be extremely unlikely that an as yet unidentified artefact would have all of the above properties.

Vance and de Kruijff make the sweeping claim (without providing any evidence) that "... we are not convinced that methylation of phosphatidylethanolamine is related to any physiological response or change in membrane fluidity". Experiments have shown a dose-dependent increase of phospholipid methylation, with the methyl donor, *S*-adenosylmethionine, which is associated with a decrease in membrane viscosity of red blood cells¹². When the methyltransferase inhibitor, *S*-adenosylhomocysteine is added with *S*-adenosylmethionine, changes in membrane viscosity are abolished with a concomitant inhibition of phospholipid methylation. It has been previously shown that extremely low concentrations of prostaglandins (a few molecules per cell) cause changes in membrane viscosity¹³. Therefore, the fact that small amounts of phospholipid are methylated is no indication that phospholipid methylation is not involved. The amounts of phospholipids methylated after the stimulation of β -adrenergic receptors in reticulocytes¹⁴ and C₆ astrocytoma¹⁰ are calculated to be 30 molecules per receptor molecule, a value which is almost equivalent to that of the annulus lipids around Ca²⁺-ATPase, a typical transmembrane protein¹⁵. Such changes in lipid composition may be sufficient to change the fluidity and to enhance the coupling of receptor to adenylate cyclase. As 1,6-diphenyl-1,3,5-hexatriene, a fluorescence probe we used for measuring membrane fluidity, is preferentially embedded in the boundary of lipid domains¹⁶, local and bulk changes of membrane fluidity cannot be separated by measuring its anisotropy. The order of potency for several agonists to increase methylation is similar to their ability to generate cyclic AMP^{10,14}. Increased methylation could be blocked by β -adrenergic antagonists. These observations are consistent with the previous reports that increasing membrane fluidity with vaccenic acid facilitates coupling of receptor to adenylate cyclase¹⁷ and that decreasing fluidity with cholesterol reduces the coupling¹⁸. Almost all the changes in phospholipid methylation reported were shown to be receptor mediated (β -adrenergic^{10,14}, concanavalin A^{5,11}, chemotactic peptide⁹, benzodiazepine¹⁰ and IgE receptors^{7,10}). Because receptors occupy only a tiny fraction of the cell surface, it would require only small amounts of methylating enzymes clustered near the receptor to produce changes in phospholipid composition. Note that phosphatidylcholine formed by transmethylation contains the major portion of arachidonic acid although transmethylation is only a minor pathway for the phosphatidylcholine synthesis compared with the CDP-choline pathway^{19,20}. Arachidonic acid

(a fatty acid of great importance for cell function) is liberated by receptor stimulation^{8,9,11}, suggesting a close coupling of receptor, methyltransferases and phospholipases and a rapid turnover of methylated phospholipids.

When cells are labelled with [$Me-^3H$]methionine, the methyl group is mainly incorporated into the phospholipid fraction via *S*-adenosylmethionine after receptor stimulation⁷. The subsequent physiological responses such as lymphocyte mitogenesis¹¹, histamine release from mast cells^{5,7} and basophils⁸

and leukocyte chemotaxis⁹ can be inhibited by treatment with 3-deazaadenosine and its analogues, which are intracellularly transformed to *S*-adenosylhomocysteine analogues, inhibitors of transmethylation reactions²¹. By treating cells with 3-deazaadenosine and its analogues, phospholipid methylation but not the transmethylation of nucleotides and proteins is inhibited, in the conditions used by us. All these experiments show the close association between receptor-mediated phospholipid methylation and physiological responses.

1. Hirata, F., Viveros, O. H., Diliberto, E. Jr & Axelrod, J. *Proc. natn. Acad. Sci. U.S.A.* **75**, 1718–1720 (1978).
2. Blustajn, J. K., Ziesel, S. H. & Wurtman, R. J. *Brain Res.* **179**, 319–327 (1979).
3. Hirata, F. & Axelrod, J. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2348–2352 (1978).
4. Crews, F. T., Hirata, F. & Axelrod, J. *J. Neurochem.* **34**, 1491–1498 (1980).
5. Hirata, F., Axelrod, J. & Crews, F. T. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4813–4816 (1979).
6. Crews, F. T., Hirata, F. & Axelrod, J. *Neurochem. Res.* (in the press).
7. Ishizaka, T., Hirata, F., Ishizaka, K. & Axelrod, J. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1903–1906 (1980).
8. Crews, F. T., Morita, Y., Hirata, F., Axelrod, J. & Siraganian, R. P. *Biochem. biophys. Res. Commun.* **93**, 42–49 (1980).
9. Hirata, F., Corcoran, B. A., Venkatasubramanian, K., Schiffmann, E. & Axelrod, J. *Proc. natn. Acad. Sci. U.S.A.* **76**, 2640–2643 (1979).
10. Strittmatter, W. J. *et al. Nature* **282**, 857–859 (1979).
11. Hirata, F., Toyoshima, S., Axelrod, J. & Waxdal, M. J. *Proc. natn. Acad. Sci. U.S.A.* **77**, 862–865 (1980).

12. Hirata, F. & Axelrod, J. *Nature* **275**, 219–220 (1978).
13. Kury, P. G., Ramwell, P. W. & McConnell, H. M. *Biochem. biophys. Res. Commun.* **56**, 178–185 (1974).
14. Hirata, F., Strittmatter, W. J. & Axelrod, J. *Proc. natn. Acad. Sci. U.S.A.* **76**, 368–372 (1978).
15. Bennett, J. P., McGill, K. A. & Warren, G. B. *Nature* **274**, 823–825 (1978).
16. Klausner, R. D., Kleinfeld, A. M., Hoover, R. L. & Karnovsky, M. J. *J. biol. Chem.* **255**, 1286–1295 (1980).
17. Rimón, G., Hanski, E., Braun, S. & Levitzki, A. *Nature* **276**, 394–398 (1978).
18. Klein, L., Moore, L. & Pastan, I. *Biochim. biophys. Acta* **506**, 42–53 (1978).
19. Ketyal, S. L. & Lombardi, B. *Lipids* **11**, 513–516 (1976).
20. Trewella, M. A. & Collins, F. D. *Biochim. biophys. Acta* **296**, 51–61 (1973).
21. Cantoni, G. L., Richards, H. H. & Chiang, P. K. in *Transmethylation* (eds Usdin, E., Borchard, R. T. & Creveling, G. R.) 155–165 (Elsevier, Amsterdam, 1979).

Reduced somatostatin-like immunoreactivity in cerebral cortex from cases of Alzheimer disease and Alzheimer senile dementia

Peter Davies*, Robert Katzman† & Robert D. Terry*

Departments of *Pathology and of Neuroscience, and of †Neurology and of Neuroscience, Albert Einstein College of Medicine, 1300 Morris Park Avenue, The Bronx, New York 10461

Both Alzheimer's disease and senile dementia of the Alzheimer type (AD/SDAT) are progressive dementias characterized neuropathologically by the presence in the cerebral cortex of numerous neurofibrillary tangles and neuritic plaques¹. We use the abbreviation AD/SDAT to denote all such cases, irrespective of age of onset². Studies of neurotransmitter-related parameters in autopsied brain tissues from patients with AD/SDAT have, to date, been confined to five putative transmitter systems. Acetylcholine-releasing neurones seem to be most markedly and consistently affected, as judged by the extensive reductions in choline acetyltransferase (ChAT) and acetylcholinesterase activities that have been reported^{3–5}. Despite numerous studies, there is no consistent evidence for the involvement of neurones releasing dopamine, noradrenaline, serotonin, or γ -aminobutyric acid in AD/SDAT⁶, nor for loss of muscarinic cholinergic receptors⁷. Thus, the involvement of cholinergic neurones in AD/SDAT seems to be specific. However, the possible involvement of neurones using other chemicals as transmitters has yet to be explored. The recent recognition of the existence of so-called 'peptidergic neurones' in the mammalian brain (for review see ref. 8) and the availability of radioimmunoassay (RIA) techniques for studying these peptides, have led us to begin a systematic investigation of neuropeptides in autopsied brain tissue from cases of AD/SDAT, and from neurologically normal individuals. We report here results obtained with a RIA for somatostatin, showing that somatostatin-like immunoreactivity in the cerebral cortex is reduced in tissue from AD/SDAT patients.

Synthetic somatostatin (Peninsula) was coupled to bovine serum albumin as described by Arimura *et al.*⁹, and an antiserum was raised in rabbits. The serum was used in RIA at a final dilution of 1:48,000 with ¹²⁵I-labelled tyrl-somatostatin as tracer (NEN; 2,200 mCi mmol⁻¹). The presence of 1 μ g per

tube of substance P, insulin, proinsulin, glucagon, Met-enkephalin, thyrotropin releasing hormone, gastrin, cholecystokinin octapeptide or neurotensin did not interfere with the assay. The limit of sensitivity of the assay was about 5 pg of synthetic somatostatin, the intra-assay variability was $\pm 7\%$ and inter-assay variation $\pm 12\%$.

Samples of right cerebral cortex were obtained at autopsy from 12 neurologically normal individuals (mean age 74 yr, range 40–101 yr) within 24 h of death (mean delay death to autopsy 11 h, range 3–24 h). The areas studied were mid-frontal gyrus, superior temporal gyrus, inferior parietal cortex and hippocampus. Neuropathological examination of left cerebral hemispheres revealed no gross or microscopic abnormalities. Similar samples were obtained from 12 cases of AD/SDAT (mean age 74 yr, range 49–92 yr; mean delay 10 h, range 2–16 h). Neuropathologically, all 12 AD/SDAT cases showed many neurofibrillary tangles and neuritic plaques in sections of frontal and parietal cortex, and in hippocampus, with no evidence of cerebrovascular disease, Parkinson's disease or other abnormality.

Extracts of brain tissues were made by boiling in 2 M acetic acid, homogenization and centrifugation at 10,000g for 30 min. Aliquots of supernatants were freeze-dried and residues were re-dissolved in phosphate-buffered saline, pH 8.2, containing 0.1% bovine serum albumin. The efficiency of extraction of ¹²⁵I-labelled somatostatin added to frozen tissue samples was 89%. The slopes of displacement curves produced by the addition of serial dilutions of extracts of brain regions were not significantly different from those produced by addition of various amounts of synthetic somatostatin. Although this indicates immunochemical similarity between material in brain extracts and synthetic somatostatin, it does not prove identity. Our preliminary studies with gel filtration have so far failed to

Table 1 Choline acetyltransferase activity in cerebral cortex samples from normal and AD/SDAT cases

	Normal			AD/SDAT		
	Means.e.m.	n	Range	Means.e.m.	n	Range
Hippocampus	575 $\pm$ 77	11	203–960	168 $\pm$ 31	10	58–367
Frontal cortex	394 $\pm$ 59	12	91–751	48 $\pm$ 9	11	7–112
Parietal cortex	305 $\pm$ 49	12	112–681	35 $\pm$ 8	12	7–77
Superior temporal gyrus	428 $\pm$ 69	12	97–673	49 $\pm$ 16	12	9–164

Choline acetyltransferase activity is expressed as nmol acetylcholine produced per h per 100 mg protein. Values for all regions of AD/SDAT cases are significantly lower than those of normals ($P < 0.01$).

Table 2 Somatostatin-like immunoreactivity in cerebral cortex samples from normal and AD/SDAT cases

	Normal			AD/SDAT		
	Means.e.m.	n	Range	Means.e.m.	n	Range
Hippocampus	3.56±0.40	10	0.84–5.35	0.95±0.04	8	0.77–1.10
Frontal cortex	2.99±0.39	12	1.11–5.79	0.72±0.15	12	0.17–1.45
Parietal cortex	3.28±0.70	12	1.27–8.89	1.47±0.40	12	0.07–3.83
Superior temporal gyrus	3.05±0.39	12	0.75–5.04	0.97±0.30	12	0.00–2.86

Somatostatin-like immunoreactivity is expressed as ng per mg protein. Values for hippocampus, frontal cortex and superior temporal gyrus are significantly lower in AD/SDAT than in normal tissue ($P < 0.01$), as are those in parietal cortex ($P < 0.05$).

reveal evidence for the presence of high molecular weight (>3,000) putative precursors of somatostatin in acetic acid extracts of human cerebral cortex. We therefore refer to immunoreactive material in human brain as somatostatin-like immunoreactivity. Choline acetyltransferase activities from these cases were determined on separate samples of the same brain regions by methods previously described³.

In agreement with our own and other published data, all four regions of AD/SDAT brains investigated showed marked, statistically significant reductions in the activity of ChAT (see Table 1). The reductions in ChAT reported here, and by one of us previously, are larger than those reported by others. One reason for this may be the ages of the patients chosen as controls, as we have discussed elsewhere³. A second factor may be our selection of only severe cases of AD/SDAT; that is, of only those cases with many plaques and tangles in both the cerebral cortex and hippocampus. The amounts of somatostatin-like immunoreactivity in all four regions of AD/SDAT brains were also significantly below those found in brains from normal individuals (Table 2). In the hippocampus, both ChAT activity and somatostatin-like immunoreactivity were reduced by about the same extent, 70.8% and 73.3%, respectively. It is possible that these two parameters are within the same neuronal structure: co-existence of peptide and amine transmitters in a single neurone has been demonstrated in other systems⁸. However, there is no other evidence for the presence of somatostatin in cholinergic terminals in the hippocampus. In the cortical regions studied, there is a greater loss of ChAT activity than of somatostatin-like immunoreactivity. The greatest loss of somatostatin-like immunoreactivity seems to occur in frontal cortex (75.9%), and the least in parietal cortex (55.2%). In both these regions the ChAT activity is reduced by almost 90%.

We have considered the possibility that the reduced somatostatin-like immunoreactivity found in the AD/SDAT cases might result from factors other than the disease. However, we have been unable to find any evidence of correlation between the amount of somatostatin-like immunoreactivity and the age of the patient, either in the control or AD/SDAT groups. Similarly, there was no correlation between the somatostatin-like immunoreactivity concentration and the delay period between death and autopsy. Because of the large number of different drug treatments administered to both the controls and the AD/SDAT cases before death, we have been unable to determine whether any one such treatment might be of particular significance.

Thus, our data seem to indicate the involvement of a second putative neurotransmitter system in AD/SDAT, and provide the first evidence for loss of a neuronal peptide in AD/SDAT cases. Efforts must now be directed towards characterizing the immunoreactive material in both normal and AD/SDAT brain. We do not yet know whether or not the loss of somatostatin-like immunoreactivity has functional significance to the patient with AD/SDAT. Further clinical and biochemical work is needed before this most important point can be clarified.

Our thanks to Aurora Thompson for technical assistance. This work was supported in part by NIH grants AG-01066, NS-03356 and NS-02255, and by a gift from A. and B. Condiolti.

Received 16 May; accepted 15 October 1980.

1. Wisniewski, H. M. & Terry, R. D. in *Neurobiology of Aging* (eds Terry, R. D. & Gershon, S.) 265–280 (Raven, New York, 1976).
2. Katzman, R. *Archs Neurol., Chicago* **33**, 217 (1976).
3. Davies, P. *Brain Res.* **171**, 319 (1979).
4. White, P. *et al. Lancet* **i**, 668 (1977).
5. Perry, E. K., Perry, R. H., Blessed, G. & Tomlinson, B. E. *Lancet* **i**, 189 (1977).
6. Terry, R. D. & Davies, P. A. *Rev. Neurosci.* **3**, 77 (1980).
7. Davies, P. & Verth, A. H. *Brain Res.* **138**, 385 (1977).
8. Hokfelt, T., Johansson, O., Ljungdahl, A., Lundberg, J. M. & Schultzberg, M. *Nature* **284**, 515 (1980).
9. Arimura, A., Sato, H., Coy, D. H. & Schally, A. V. *Proc. Soc. exp. Biol. Med.* **148**, 784 (1975).

A new class of angiotensin-converting enzyme inhibitors

A. A. Patchett, E. Harris, E. W. Tristram, M. J. Wyvratt, M. T. Wu, D. Taub, E. R. Peterson, T. J. Ikeler, J. ten Broeke, L. G. Payne, D. L. Ondeyka, E. D. Thorsett, W. J. Greenlee, N. S. Lohr, R. D. Hoffsommer, H. Joshua, W. V. Ruyle, J. W. Rothrock, S. D. Aster, A. L. Maycock, F. M. Robinson & R. Hirschmann

New Lead Discovery Department, Basic Research, Merck Sharp and Dohme Research Laboratories, Division of Merck and Co. Inc., Rahway, New Jersey 07065

C. S. Sweet, E. H. Ulm, D. M. Gross, T. C. Vassil & C. A. Stone

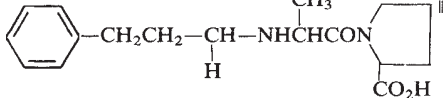
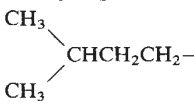
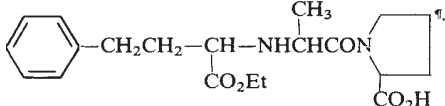
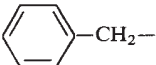
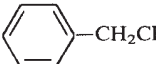
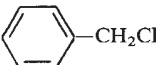
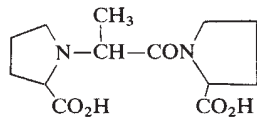
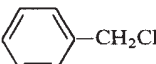
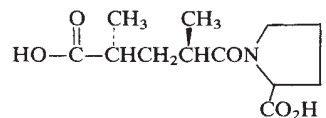
Merck Institute for Therapeutic Research, West Point, Pennsylvania 19486

Much current attention focuses on the renin-angiotensin system in relation to mechanisms controlling blood pressure and renal function. Recent demonstrations (ref 1, ref. 2 and refs therein) that angiotensin-converting enzyme inhibitors show promising clinical antihypertensive properties have been of particular interest. We now report on the design of a novel series of substituted *N*-carboxymethyl-dipeptides which are active in inhibiting angiotensin-converting enzyme at nanomolar levels. We suggest that these compounds are transition-state inhibitors and that extensions of this design to other metalloendopeptidases merit further study.

The biological profile of three of these inhibitors was assessed in rats and dogs by the antagonism of the pressor response to angiotensin I. High converting-enzyme inhibitory activities (at doses as low as 2.3 µg per kg) were demonstrated after intravenous (i.v.) administration. Given orally (p.o.) in dogs these three compounds showed good activities with a duration of action of at least 6 h at doses of 0.3 mg per kg. This particularly long duration of action and their apparent specificity may make these compounds useful in the treatment of hypertension.

Angiotensin-converting enzyme (peptidyl-dipeptide hydrolase, EC 3.4.15.1) generates the powerful vasoconstrictor substance angiotensin II by removing the C-terminal dipeptide from the precursor decapeptide angiotensin I (ref. 3). The enzyme also inactivates the vasodilating substance bradykinin⁴. That angiotensin-converting enzyme inhibitors might display useful antihypertensive properties has been demonstrated clinically by the nonapeptide SQ 20,881 (ref. 1) and more recently by many studies² on the orally active angiotensin-converting enzyme inhibitor captopril (SQ 14,225)⁵. The latter structure, D-3-mercapto-2-methylpropanoyl-L-proline (resulting from work by the Squibb group) inhibits the enzyme with a reported IC₅₀ of 2.3×10^{-8} M (ref. 6).

Table 1 *In vitro* converting-enzyme inhibitory activities of *N*-carboxymethyl-L-alanyl-L-proline, derivatives and analogues*

R₁ CH-Ala-Pro CO₂H					
No.	R ₁	IC ₅₀	No.	R ₁	IC ₅₀
1	H—†	2.4 × 10 ⁻⁶ M	Related analogues		
2	CH ₃ —	9 × 10 ⁻⁸ M	9		5.8 × 10 ⁻⁶ M
3	CH ₃ CH ₂ —	1.7 × 10 ⁻⁸ M			
4		2.6 × 10 ⁻⁹ M	10		1.2 × 10 ⁻⁶ M
5		3.9 × 10 ⁻⁸ M			
6		3.8 × 10 ⁻⁹ M	Maleate salt (S, S, S configuration)		
6a	 (S configuration)	1.2 × 10 ⁻⁹ M	11		1.3 × 10 ⁻⁶ M
6b	 (R configuration)	8.2 × 10 ⁻⁷ M			
7	NH ₂ —(CH ₂) ₄ —§	2.2 × 10 ⁻⁹ M	12		4.8 × 10 ⁻⁶ M
8	NH ₂ —(CH ₂) ₅ —	5.2 × 10 ⁻⁹ M			

Converting enzyme was prepared from fresh citrated hog plasma by the method of Dorer *et al.*¹⁷ up to fraction 'B', followed by chromatography on Sephadex G-200 and collection of the fractions of highest specific activity. This enzyme was chloride-dependent, 90% inhibited by EDTA at 1 × 10⁻⁴ M and yielded His-Leu from angiotensin I as determined by TLC. Assays adapted from the method of Piquilloud *et al.*¹⁸ were conducted in the following manner. Test compounds were dissolved at 5 × 10⁻³ M in 0.05 M borate-phosphate buffer, pH 8.0, 0.17 M in NaCl. Further dilutions were made with this buffer. To 100 μl of each test-compound solution was added 100 μl each of 3.7 × 10⁻⁴ M substrate solution Cbz-Phe-His-[¹⁴C-Leu]-OH, 5,500 c.p.m. per tube and converting-enzyme solution diluted from stock so as to give approximately 1,200 c.p.m. in the control supernatant after incubation. The final concentration of substrate was 1.23 × 10⁻⁴ M. *K_m* in this assay system was between 1 and 2 × 10⁻⁴ M. The reaction was incubated 1 h at 30 °C, then terminated by the addition of 1 ml of a 40% (settled solids) suspension in 0.05 M borate buffer of benzoylated DEAE (BD) cellulose (Schwarz-Mann 100–200 mesh, defined), pH 11 or 20% BD Sephadex (Biorad) in 0.05 M borate buffer, pH 7.7. Unreacted substrate was bound to the BD adsorbant and separated from supernatant by filtration through a cotton-plugged Sarpette (size 102/2, Walter Sarstedt). The entire supernatant plus a 0.5-ml buffer wash of the BD residue was collected in a scintillation vial. Counting was done in 10 ml PCS (Amersham Searle) for 10 min. Minus-enzyme blanks were run through the assay and subtracted from each control and test count. IC₅₀ values (concentration for 50% inhibition) were determined by linear regression of logit against log concentration over a 20–80% inhibition range.

* Except where indicated, compounds were synthesized by the reductive condensation of α-ketoacids or esters with dipeptides and were assayed without the separation of diastereomers. All compounds were characterized by satisfactory mass spectra and NMR data and by TLC on silica-gel plates.

† This compound was prepared by the alkylation of L-alanyl-L-proline in aqueous solution at 85 °C by the portion-wise addition of chloroacetic acid over 1 h with maintenance of pH 8–9.

‡ Diastereomers were separated by chromatography on 200–400 mesh XAD-2 polystyrene resin (Rohm and Haas) using 95:5 (v/v) 0.1 M aqueous NH₄OH-methanol and were crystallized from water. Compound 6a (eluted first) [α]_D²⁵ = –67.0 °C (0.1 M HCl); compound 6b (eluted second) [α]_D²⁵ = –101.6 °C (0.1 M HCl). Compound 6a: Analyses calculated for C₁₈H₂₄N₂O₅·1/2H₂O: C, 60.40; H, 7.05; N, 7.84. Found: C, 60.60; H, 7.01; N, 7.79.

§ Compound 7 was prepared from ε-benzoyloxycarbonyl-L-lysine ethyl ester and *N*-pyruvoyl-L-proline in dry tetrahydrofuran. Sodium cyanoborohydride was added to this mixture after it had been stirred at room temperature for 4 h with powdered no. 4A molecular sieves. Protecting groups were removed by saponification of the ethyl ester with 0.1 M NaOH overnight followed by hydrogenolysis of the benzoyloxycarbonyl group in 1:1 ethanol-water over 10% Pd/C catalyst at 40 p.s.i. Diastereomers were not separated.

|| Hydrocinnamaldehyde was reductively condensed with L-alanyl-L-proline using sodium cyanoborohydride in aqueous ethanol.

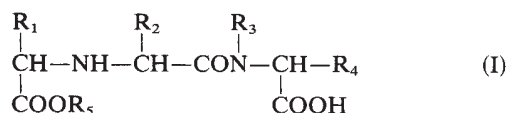
¶ This salt was precipitated from an acetonitrile solution of *N*-(1-ethoxycarbonyl-3-phenylpropyl)-L-alanyl-L-proline (mixed diastereomers) using a half-molar equivalent of maleic acid. Repeated recrystallization from acetonitrile afforded pure compound 10 [α]_D²⁵ = –42° (CH₃OH). Analyses calculated for C₂₄H₃₂N₂: C, 58.53; H, 6.55; N, 5.69. Found: C, 58.49; H, 6.53; N, 5.61. As even traces of 6a in this sample would give significant inhibition, the IC₅₀ may overestimate its potency.

Compound 10 is also designated MK-421.

** Compound 11 was prepared from *N*-pyruvoyl-L-proline and L-proline benzyl ester using sodium cyanoborohydride in methanol. The benzyl group was subsequently removed by hydrogenolysis using 10% Pd/C in aqueous ethanol at 40 p.s.i. Diastereomers were not separated.

†† Compound 12 was prepared from d1-2,4-dimethylglutaric acid¹⁹ which as its anhydride was reacted with L-proline benzyl ester in CH₂Cl₂. The benzyl group was removed by hydrogenolysis and the diastereomers separated by successive chromatography over LH-20 using CH₃OH followed by HPLC on a Whatman Partisil-10 ODS-2 preparative column using H₂O:CH₃CN:HOAc (75:25:1).

The most common side effects accompanying the clinical use of captopril are rashes and loss of taste both of which usually clear on withdrawal or reduction of dose². As similar side effects are observed with penicillamine⁷, we looked for potent specific inhibitors of angiotensin-converting enzyme lacking a mercapto function and characterized by weak chelating properties. We report here the design of such compounds illustrated by generic formula (I).



These substituted *N*-carboxymethyl-dipeptides are conveniently prepared by the reductive coupling of α-ketoacids or esters to the amino terminus of suitably protected dipeptides utilizing sodium cyanoborohydride or hydrogenation with palladium-on-carbon or Raney nickel. Typically, 1.42 g of 2-oxo-4-phenylbutyric acid was dissolved in 5 ml of water by adjusting the pH to 6.8 with 50% NaOH. L-Alanyl-L-proline (0.27 g) was then added and the solution treated with 0.28 g of sodium cyanoborohydride in 5 ml of water. After stirring at room temperature overnight, 35 ml of Dowex 50 X-2, 50–100 mesh on the acid cycle, was added to absorb the product and destroy any residual sodium cyanoborohydride. The slurry was then added to a column containing 20 ml of Dowex 50 and

$$\text{C}_6\text{H}_5-\text{CH}_2\text{CH}_2-\underset{\text{CO}_2\text{H}}{\text{CH}}-\text{A}_1-\text{A}_2$$

transition state-like geometry is attained at the scissile CONH locus via the CHCO_2H and NH groups. Cumulative binding contributions are provided by the CO_2H , NH and R_1 groups as indicated in the analogue studies. It remains to be determined if the NH is protonated while bound to the enzyme, thus contributing a potential electrostatic interaction.

These compounds were evaluated for their ability to antagonize the pressor effect of angiotensin I administered i.v. to rats and dogs. Doses of angiotensin I were selected which represented the midpoint of their respective dose-response curves. In anaesthetized rats this was 100 ng per kg i.v., and it was 600 ng per kg i.v. in anaesthetized dogs. In conscious dogs, angiotensin I was administered at 150 ng per kg i.v. Data for three of our inhibitors are summarized in Table 3 and compared with captopril. Clearly compounds 6a, 10 and 21a were highly active by the i.v. route in rats. The same was true in dogs except for compound 10 whose ester functionality was found to be hydrolysed by dog liver homogenates but poorly so by dog plasma. Hydrolysis of compound 10 in rat plasma was rapid. (Hydrolysis of ^{14}C -Proline UL) compound 10 to compound 6a was estimated by incubation of 10^{-5} M compound 10 with undiluted plasma or a 1:5 homogenate of liver in 5 mM NaHCO_3 . The reaction was stopped at various times by acidification with HCl and after isolation using XAD-4 resin the proportion of compounds 6a to 10 was determined by TLC (Quantum LQDF developed with $n\text{-BuOH}:\text{HOAc}:\text{H}_2\text{O}$, 4:1:1) and by the degree of inhibition of angiotensin-converting enzyme.)

All three compounds were orally active in rats and dogs. Comparative p.o. converting-enzyme inhibitory activities in dog are shown in Fig. 1. Compounds 6a, 10 and 21a displayed high potency in inhibiting the blood pressure response to angiotensin I and all had a long duration of action. Detailed pharmacological studies further demonstrating the p.o. and i.v. activities of some of these compounds and their antihypertensive properties will be reported elsewhere^{14,15}.

Clinical studies now in progress¹⁶ with compound 10 (MK-421) should help to determine if converting-enzyme inhibitors lacking a mercapto function have advantages in respect to side effects². More importantly, we believe that our angiotensin-converting enzyme inhibitors have important design implications for the synthesis of inhibitors of other metalloprotease carboxypeptidases and possibly of metalloendopeptidases more generally. In our design a CO_2H , an NH and an R group are arranged in tetrahedral geometry at the locus of the scissile amide group. Their binding energies in concert replace the Zn^{2+} /sulphydryl interaction which is a required feature of captopril^{3,6,8}.

We thank Drs C. Hoffman, S. Michelson and H. Schwam for isolations of angiotensin-converting enzyme, and Dr J. A. McCauley for determination of the zinc metal stability constant.

Received 14 May; accepted 18 September 1980.

- Gavras, H. *et al.* *New Engl. J. Med.* **291**, 817–821 (1974).
- Atkinson, A. B. & Robertson, J. I. S. *Lancet* **ii**, 836–839 (1979).
- Skeggs, L. T., Marsh, W. H., Kahn, J. R. & Shumway, N. P. *J. exp. Med.* **99**, 275–282 (1954).
- Yang, H. Y. T., Erdős, E. G. & Levin, Y. *Biochim. biophys. Acta* **214**, 374–376 (1970).
- Ondetti, M. A., Rubin, B. & Cushman, D. W. *Science* **196**, 441–444 (1977).
- Cushman, D. W., Cheung, H. S., Sabo, E. F. & Ondetti, M. A. *Biochemistry* **16**, 5484–5491 (1977).
- Levine, W. G. in *The Pharmacological Basis of Therapeutics* 5th edn (eds Goodman, L. S. & Gilman, A.) 919–920 (Macmillan, New York, 1975).
- Cushman, D. W., Cheung, H. S., Sabo, E. F. & Ondetti, M. A. *Fedn Proc.* **33**, 2778–2782 (1979).
- Soffer, R. L. A. *Rev. Biochem.* **45**, 73–94 (1976).
- Cheung, H. S., Wang, F. L., Ondetti, M. A., Sabo, E. F. & Cushman, D. W. *J. biol. Chem.* **255**, 401–407 (1980).
- Beckner, C. F. & Caprioli, R. M. *Biochem. biophys. Res. Commun.* **93**, 1290–1296 (1980).
- Stability Constants of Metal Ion Complexes* 520 (Spec. Publ. no. 17, The Chemical Society, London, 1964).
- Wolfenden, R. A. *Rev. Biophys. Bioengng* **5**, 271–306 (1976).
- Sweet, C. S. *et al.* *J. Pharmac. exp. Ther.* (submitted).
- Gross, D. M. *et al.* *J. Pharmac. exp. Ther.* (submitted).
- Biollaz, J. *et al.* *Clin. Pharmac. Ther.* (submitted).
- Dorer, F. E., Kahn, J. R., Lentz, K. E., Levine, M. & Skeggs, L. T. *Circulation Res.* **34**, 824–827 (1974).
- Piquilloud, Y., Reinharz, A. & Roth, M. *Biochim. biophys. Acta* **206**, 136–142 (1970).
- Allinger, N. L. *J. Am. chem. Soc.* **81**, 232 (1959).
- Rubin, B. *et al.* *J. Pharmac. exp. Ther.* **204**, 271 (1978).

³H-apomorphine labels both dopamine postsynaptic receptors and autoreceptors

Pierre Sokoloff, Marie-Pascale Martres & Jean-Charles Schwartz

Unité 109 de Neurobiologie, Centre Paul Broca de l'INSERM, 2ter, rue d'Alésia, 75014 Paris, France

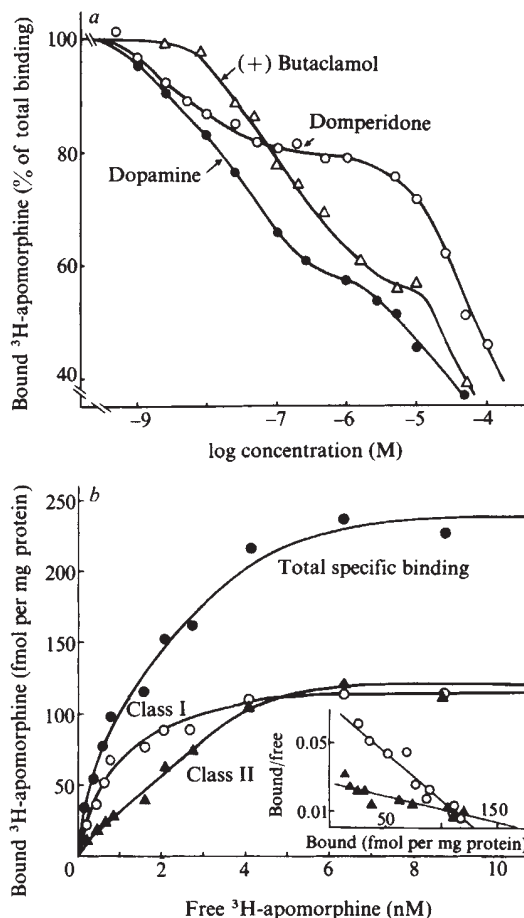
Evidence points to the existence of multiple classes of dopamine (DA) receptor in mammalian brain which can be distinguished by their pharmacological specificities and localizations, and by the actions they mediate¹. Among them, dopaminergic autoreceptors are regulatory receptors presumed to be present in the membrane of the DA neurones themselves, and believed to mediate an inhibition of these neurones' activity, either at nerve endings or on cell bodies^{2–5}. However, the pharmacology of autoreceptors remains to be established because attempts to characterize autoreceptors by ³H-ligand binding techniques have produced controversial data. Thus Seeman and co-workers stated that ³H-apomorphine selectively labels autoreceptors^{6,7}, whereas Creese *et al.* concluded that this ligand selectively labels postsynaptic DA receptors^{8,9}. In addition, large differences in the capacity and drug specificity of ³H-apomorphine receptor sites in rat striatum have been reported^{10–13}. We demonstrate here that ³H-apomorphine labels two classes of DA receptor, distinguishable using domperidone, a selective DA antagonist^{14–17}. Lesion studies indicate that they correspond to a certain class of postsynaptic receptor and to autoreceptors, respectively. Each of these classes displays a clearly distinct pharmacological specificity for antipsychotics and DA agonists.

Total binding of ³H-apomorphine was studied at 6 nM in the presence of increasing concentrations of DA, (+)butaclamol (an antipsychotic drug) or domperidone. As shown in Fig. 1a, DA and (+)butaclamol displayed monophasic inhibition curves up to concentrations of 1 and 10 μM respectively. At these concentrations, a small but reproducible plateau was observed allowing definition of the specific binding (45% of total binding) in conditions similar to those used by others^{6,8,10,18}. In contrast, domperidone inhibited this specific binding in a clearly biphasic manner, and a wide plateau appeared between approximately 50 nM and 2 μM . This allowed definition of two pharmacologically distinct classes of ³H-apomorphine binding site: class I, comprising sites recognized by low concentrations of domperidone ($\text{IC}_{50} = 5\text{ nM}$), and class II, composed of sites poorly recognized by domperidone ($\text{IC}_{50} \sim 10\text{ }\mu\text{M}$).

Saturation of these two classes of binding site by ³H-apomorphine in increasing concentrations was studied separately—class I sites by determining the fraction of binding inhibited by 200 nM domperidone, and class II sites by determining the binding which occurred in the presence of 200 nM domperidone and was inhibited by 1 μM dopamine (Fig. 1b). Both classes of binding site distinguished by domperidone are characterized by high affinities for ³H-apomorphine (K_d 0.9 and 2.6 nM), and the capacity of each one amounts to about 120 fmol per mg protein (7 pmol per g tissue).

Lesion data show that these two classes of site are located on different populations of striatal neurones (Table 1). Class I sites seem to be mostly, if not exclusively, present on intrinsic neurones, as the drop in their binding capacity after kainate lesion (–57%) is the same as the reduction occurring in the activity of glutamate decarboxylase, a marker of striatal γ -aminobutyric acid neurones. In addition, the number of class I sites is significantly enhanced after 6-hydroxydopamine(6-OHDA)-induced degeneration of afferent DA

Fig. 1 ^3H -Apomorphine binding to rat striatal particulate fraction. The dissected tissues (40–50 mg) from Sprague-Dawley rats (C. River, 200–250 g) were gently sonicated (Ultrasons-Annemasse; sonotrode $d = 4.5$ mm) at 4°C in 2 ml 50 mM Tris-HCl buffer pH 7.4 (in 15 ml borosilicated glass tubes). After a three-fold dilution, the homogenate was centrifuged (5,000 g min) and the supernatant diluted two-fold and centrifuged (200,000 g min) in 12 ml cellulose nitrate test tubes. The pellet was resuspended in 2 ml Tris-HCl buffer using a glass-glass homogenizer (Dounce, clearance 70–120 μm) and incubated for 12 min at 37°C in a metabolic shaker (140 c min $^{-1}$). After fivefold dilution and centrifugation (200,000 g min), the particulate fraction was resuspended with the Dounce homogenizer in 2 ml freshly prepared 50 mM Tris-HCl buffer pH 7.4 containing 120 mM NaCl, 5 mM KCl, 1 mM CaCl_2 , 1 mM MgCl_2 , 0.1% ascorbic acid and 10 μM pargyline ('Tris-ions buffer'). The suspension was briefly sonicated (4 s) before use. In preliminary experiments, we checked that this procedure for tissue preparation led to an almost complete elimination of ^3H -DA added before homogenization. In the binding assay, 200 μl freshly prepared tissue suspension were preincubated for 20 min at 30°C in polystyrene test tubes. Incubations were initiated by addition of 200 μl Tris-ions buffer containing [8, 9- ^3H]apomorphine (30.6 Ci mmol $^{-1}$, NEN) and various drugs, and were carried on for 30 min at 30°C . Drugs were dissolved in either Tris-ions buffer or 0.01–0.05% lactic acid and diluted into glass tubes with Tris-ions buffer. Lactic acid at the final concentration used did not interfere with the binding. Incubations were stopped by dilution with 3 ml ice-cold Tris-ions buffer followed by rapid filtration under vacuum (0.2–0.4 bar) through Whatman GF/B filters. Tubes were rinsed with 3 ml ice-cold Tris-ions buffer and the filters were washed with 2×5 ml of the same slowly delivered buffer. After drying the filters, the radioactivity was counted by liquid scintillation spectrometry in 12 ml toluene-Triton X-100-PP-POPOP (21.11–16.5 g–0.45 g) at an efficiency of 45%. Binding on filters represented less than 0.5% of the total radioactivity. *a*, Inhibition by three dopaminergic agents. Striatal particulate fractions were incubated in the presence of 6 nM ^3H -apomorphine and dopamine (●), (+) butaclamol (Δ) or domperidone (○) in increasing concentrations. Values are expressed in per cent of total binding (523 ± 73 fmol per mg protein) and represent the mean of 6–8 determinations performed in two separate experiments. s.e.m. of these values (not shown) were less than 3% of total binding. *b*, Saturation and Scatchard analysis (inset). Total specific binding (●) was that inhibited by 1 μM dopamine. Binding to class I sites (○) was that inhibited by 200 nM domperidone. Binding to class II sites (Δ) was that occurring in the presence of 200 nM domperidone minus that in the presence of 1 μM dopamine. At 6 nM ^3H -apomorphine, radioactivity bound to the tissue was about 3,500 c.p.m. without inhibiting drug, 2,700 c.p.m. in the presence of 200 nM domperidone and 1,900 c.p.m. in the presence of 1 μM dopamine. Total specific binding as well as binding to class I or II sites increased linearly with protein level between 0.1 and 0.3 mg per incubation. Values are the means of 3–10 determinations performed in three independent experiments. Analysis of these data by an iterative computing method based on the least squares³² gave the following parameters: $(B_{\text{max}})_I = 131 \pm 12$ fmol per mg protein, $(K_d)_I = 0.92 \pm 0.24$ nM, $(nH)_I$; (Hill constant) $= 0.95 \pm 0.17$ for the first class and $(B_{\text{max}})_{II} = 151 \pm 32$ fmol per mg protein, $(K_d)_{II} = 2.64 \pm 1.18$ nM, $(nH)_{II} = 1.17 \pm 0.30$ for the second; values are in good agreement with those obtained by Scatchard plot analysis.



neurones. On the other hand, the number of class II sites is not affected by kainate lesions but is significantly reduced after two kinds of selective lesion of DA neurones (Table 1). The changes observed after degeneration of DA neurones consist of a reduction in the number of class II sites and a rise in the number of class I sites, with no change in the affinity of domperidone for either class of site. Thus, when ^3H -apomorphine binding in striata of 6-OHDA-lesioned animals (eight rats in two separate experiments) was displaced by domperidone in increasing concentrations, a curve (not shown) similar to that in Fig. 1a was obtained. It also exhibited a large plateau whose height on the ordinate was reduced compared to the control.

The highly heterogeneous but more or less parallel distribution of the two classes of site among the various regions of rat brain reflects DA terminal distribution and thus supports the idea that DA receptors are labelled in both cases¹⁹.

The drug specificity of sites I and II was established in conditions allowing selective labelling of each class (Table 2). For class I sites, which displayed the highest affinity for ^3H -apomorphine, a low ligand concentration was used (0.8 nM). For class II sites, inhibitory potencies were established in the presence of 6 nM ^3H -apomorphine plus 200 nM domperidone to prevent labelling of class I sites. In both cases it was checked that 'nonspecific' binding (binding in the presence of 200 nM of domperidone and of 1 μM DA for class I and II sites, respectively) was not significantly inhibited by the various drugs at their IC_{50} . ^3H -Apomorphine binding to both classes of site was stereospecific and exhibited distinct pharmacological specificity, as shown by the relative affinities of several DA agonists and antagonists, which contrasted with the lack of potency of a variety of non-dopaminergic agents (Table 2). The low potency

of benztropine, an inhibitor of DA uptake, rules out the possibility that class II sites might be linked to this transport system.

Hence the two classes of site have clearly different localizations and pharmacological specificities. Class I sites consist of postsynaptic DA receptors, present on intrastriatal neurones. The increase in their number after ablation of DA neurones probably reflects denervation hypersensitivity^{8,20,21}. They are characterized by the high affinity of DA and both DA agonists and DA antagonists (K_i in nanomolar range), and thus resemble the ^3H -apomorphine binding sites described by others¹³. Their capacity is about one-third that of DA receptors labelled with ^3H -butyrophenones^{22–24} or ^3H -domperidone¹⁶. However, class I sites correspond to a subpopulation of sites constituting ~30% of the total²⁵ that have a much higher affinity for DA ($K_i = 7$ nM) than most of the ^3H -domperidone binding sites ($K_i = 2$ μM)²⁶. Class I sites cannot be D_1 receptors, that is DA receptors linked to adenylate cyclase¹, because the affinities of domperidone and sulpiride for these sites (K_i 0.5 nM and 100 nM, respectively) are several orders of magnitude higher than for the DA-sensitive cyclase of rat striatum ($K_i > 10$ μM in both cases)^{16,17,27}. They may be postsynaptic DA receptors mediating certain behavioural effects of low doses of DA agonists which are prevented by neuroleptics in moderate dosage and which have often been ascribed to the stimulation of autoreceptors^{28,29}. In accordance with the nomenclature of Keabian and Calne¹ we suggest calling these sites D_2 receptors.

On the other hand, class II sites correspond, at least in part, to DA autoreceptors, as shown by the effects of the two neurotoxins 6-OHDA and kainate (Table 1). However, the partial decrease of class II sites, in contrast to the nearly complete loss of tyrosine hydroxylase activity after 6-OHDA, does not

exclude the possibility that ~50% of these sites may be localized on cells other than DA neurones. Autoreceptors, like class I sites, but unlike the major fraction of post-synaptic receptors labelled by ^3H -butyrophenones and ^3H -domperidone, are characterized by their high affinity (K_i in nanomolar range) for most DA agonists. This agrees with electrophysiological, behavioural and biochemical observations suggesting that autoreceptors are stimulated by DA agonists at low dosage^{2,4,28,29}. An interesting exception is bromocriptine, which is approximately 60 times less potent on autoreceptors than on postsynaptic receptors labelled either with ^3H -apomorphine (class I sites) or ^3H -butyrophenones²³. Antipsychotic agents belonging to various chemical groups are recognized by autoreceptors but note that their potency is always lower than on postsynaptic receptors, that is class I sites (Table 2) or ^3H -butyrophenones and ^3H -domperidone binding sites^{16,22-24}. The number of class II sites found here is approximately five times higher than reported earlier⁶ but only slightly higher than that of the ^3H -apomorphine sites recently described by Titeler *et al.* under the name of D_3 receptors. These D_3 receptors display a drug specificity similar to that of class II sites but were not localized by lesion studies¹². Thus, class II sites could be termed D_3 receptors.

The discrepancies between various studies concerning drug specificity and localization of sites labelled by ^3H -apomorphine might be attributable to their heterogeneity with preferential labelling of either class I or class II sites in a given laboratory according to small differences in experimental conditions.

Finally, note that the differential drug specificity of autoreceptors might have therapeutic implications. Because 'on-off' reactions in the treatment of Parkinson's disease (paradoxically inhibitory effects produced by DA agonists when their plasma levels are low) are attributed to selective stimulation of autoreceptors, the low frequency of such reactions in patients treated with bromocriptine³⁰ might be explained by the low potency of this compound on autoreceptors. In addition, it is unlikely that the antipsychotic activity of antagonists is related to the block-

Table 2 Inhibition by various agents of the ^3H -apomorphine binding on the two classes of striatal sites

Agent	K_i values (nM)		$(K_i)_{II}/(K_i)_I$
	Class I	Class II	
DA antagonists			
Spiperone	0.20	522	2,610
Domperidone	0.52	3,150	6,050
(+) Butaclamol	1.7	13.5	8
α -Flupentixol	6.8	167	25
Haloperidol	7.2	332	46
Sulpiride	97.3	100,600	1,030
(-) Butaclamol	1,121	1,090	1
DA agonists			
N-Propyl-norapomorphine	0.40	3.5	9
Apomorphine	0.65	2.9	4
ADTN	1.1	4.5	4
Bromocriptine	7.2	420	58
Dopamine	9.9	6.1	0.6
Miscellaneous			
Noradrenaline	—	65	—
Phentolamine	—	6,100	—
Clonidine	—	>30,000	—
Serotonin	—	5,780	—
Morphine	—	>150,000	—
Benzotriptine	—	4,900	—

Binding on sites of class I was measured in the presence of 0.8 nM ^3H -apomorphine and defined as the excess over blank values obtained in the presence of 200 nM domperidone. Values in a typical experiment were: 650 c.p.m. (total binding), 350 c.p.m. (in the presence of 200 nM domperidone) and 250 c.p.m. (in the presence of 1 μM dopamine). Binding on sites of class II was measured in the presence of 6 nM ^3H -apomorphine and 200 nM domperidone and defined as the excess over blank values obtained in the presence of 1 μM dopamine (see Fig. 1b legend). The effect of each agent was determined at 4–6 concentrations (triplicate determinations) and IC_{50} calculated by log-probit analysis of pooled data from 2–4 separate experiments. K_i values were derived from IC_{50} according to the relationship $K_i = \text{IC}_{50}/(1 + S/K_d)$. (K_d)_I = 0.92 nM and (K_d)_{II} = 2.64 nM.

Table 1 Effect of various lesions on the capacity of the two classes of ^3H -apomorphine binding site in rat striatum

Type of lesion		^3H -Apomorphine binding (fmol per mg protein)		
		Total specific	Class I	Class II
Kainate into the striatum (7)	Contralateral	203 $\pm$ 15	120 $\pm$ 10	83 $\pm$ 14
	Ipsilateral	133 $\pm$ 14	50 $\pm$ 11	83 $\pm$ 10
	Difference	−33 $\pm$ 8%	−57 $\pm$ 9%	+3 $\pm$ 13%
6-OHDA into the medial forebrain bundle (5)	Control	249 $\pm$ 31	120 $\pm$ 16	129 $\pm$ 12
	Lesioned	219 $\pm$ 28	155 $\pm$ 24	64 $\pm$ 12
	Difference	−12%	+28%	−50%
6-OHDA into the substantia nigra (14)	Contralateral	235 $\pm$ 7	126 $\pm$ 8	109 $\pm$ 5
	Ipsilateral	220 $\pm$ 10	153 $\pm$ 9	68 $\pm$ 8
	Difference	−5 $\pm$ 4%	+24 $\pm$ 7%	−37 $\pm$ 8%

Lesions were performed in chloral-anaesthetized rats positioned in a David Kopf stereotaxic apparatus. Kainic acid (2.5 μg in 1 μl of solution titrated at pH 7.4) was unilaterally injected (2.5 min infusion) into the striatum at coordinates³³ A = 7.5 mm, L = 2.8 mm and H = 5.1 mm and animals were killed 10–15 days later. Striatal glutamate decarboxylase³⁴ was decreased by 57 $\pm$ 5% in the lesioned compared to contralateral sides (94.2 $\pm$ 2.2 nmol per h per mg protein of $^{14}\text{CO}_2$ formed). In contrast, DOPA decarboxylase activity³⁵ was not significantly affected (46.2 $\pm$ 0.9 compared to 46.3 $\pm$ 2.1 nmol per h per mg protein of $^{14}\text{CO}_2$ formed). 6-OHDA (8 μg in 4 μl of saline containing 0.1% Na ascorbate) was bilaterally injected (5-min infusion) into the medial forebrain bundle at coordinates³⁶ A = 3.2 mm, L = 1.4 mm, H = −3.0 mm and animals were killed 6–9 days later. Striatal tyrosine hydroxylase activity³⁷ was significantly decreased by 79% (0.64 $\pm$ 0.19 compared to 3.09 $\pm$ 0.21 nmol per h per mg protein of $^3\text{H}_2\text{O}$ formed). In the third group of animals, 6-OHDA (8 μg) was unilaterally injected into the substantia nigra as described³⁸ and animals killed 4–5 weeks later. Striatal tyrosine hydroxylase activity was decreased by 89 $\pm$ 3% in the lesioned compared to contralateral sides. Binding was measured in the presence of 6 nM ^3H -apomorphine in the conditions described in Fig. 1b legend. Triplicate determinations on tissues from single animals, the number of which is indicated in parentheses. Binding on the contralateral side of animals lesioned unilaterally with either kainate or 6-OHDA did not significantly differ from that on unlesioned animals (not shown).

NS, not significant. * $P < 0.01$, † $P < 0.005$, ‡ $P < 0.001$; two-tailed *t*-test. § $P < 0.05$; one-tailed *t*-test.

ade of autoreceptors as indicated, for instance, by the extremely low potency of sulpiride and by the low potency of spiperone. Furthermore, autoreceptors are probably not blocked during clinical treatment with most antipsychotics at a moderate dosage. In the treatment of psychotics it would therefore seem rational to combine DA agonists with high potency for autoreceptors, and DA antagonists with high relative affinity for postsynaptic receptors. Note that such an apparently paradoxical combination of apomorphine and antipsychotics was recently reported to improve psychotic symptoms significantly, whereas bromocriptine did not show the same activity as apomorphine³¹.

This work was supported by a grant from the CNRS (Pharmacologie des récepteurs des neuromédiateurs).

Received 24 March; accepted 17 September 1980.

1. Keabian, J. W. & Caine, D. B. *Nature* **277**, 93–96 (1979).
2. Carlsson, A. in *Pre- and Post-synaptic Receptors*, Annual ACNP Meet., Puerto-Rico (eds Usdin, E. & Bunney W. E.) (Dekker, New York, 1975).
3. Roth, R. H. *Commun. Psychopharmac.* **3**, 429–445 (1979).
4. Aghajanian, G. K. & Bunney, B. S. *Naunyn-Schmiedeberg's Arch. Pharmacol.* **297**, 1–7 (1977).
5. Glowinski, J. & Cheramy, A. in *Catecholamines: Basic and Clinical Frontiers* Vol. 1 (eds Usdin, E., Kopin, I. J. & Barchas, J.) 231–239 (Pergamon, New York, 1978).
6. Nagy, J. I., Lee, T., Seeman, P. & Fibiger, H. C. *Nature* **274**, 278–281 (1978).
7. Weinreich, P., Lee, T. & Seeman, P. *Soc. Neurosci. Abstr.* **4**, 1687 (1978).
8. Creese, I. & Snyder, S. H. *Eur. J. Pharmac.* **56**, 277–281 (1979).
9. Creese, I., Usdin, T. & Snyder, S. H. *Nature* **278**, 577–578 (1979).
10. Creese, I., Prosser, T. & Snyder, S. H. *Life Sci.* **23**, 495–500 (1978).
11. Creese, I., Stewart, K. & Snyder, S. H. *Eur. J. Pharmac.* **60**, 55–66 (1979).
12. Titeler, M., List, S. & Seeman, P. *Commun. Psychopharmac.* **3**, 411–420 (1979).
13. Leysen, J. E. *Commun. Psychopharmac.* **3**, 397–410 (1979).
14. Niemegeers, C. J. F. & Janssen, P. A. J. *Proc. 7th Int. Congr. Pharmac.* Paris, Abstr. no. 68 (1978).
15. Martres, M. P., Baudry, M. & Schwartz, J. C. *Life Sci.* **23**, 1781–1784 (1978).
16. Baudry, M., Martres, M. P. & Schwartz, J. C. *Naunyn-Schmiedeberg's Arch. Pharmac.* **308**, 231–237 (1979).
17. Watling, K. J., Dowling, J. E. & Iversen, L. L. *Nature* **281**, 578–580 (1979).
18. Thal, L., Creese, I. & Snyder, S. H. *Eur. J. Pharmac.* **49**, 295–299 (1978).
19. Sokoloff, P., Martres, M. P. & Schwartz, J. C. *Naunyn-Schmiedeberg's Arch. Pharmac.* (in the press).
20. Creese, I., Burt, D. R. & Snyder, S. H. *Science* **197**, 596–598 (1977).
21. Schwartz, J. C., Costentin, J., Martres, M. P., Pzotais, P. & Baudry, M. *Neuropharmacology* **17**, 665–685 (1978).
22. Seeman, P., Chau-Wong, M., Tedesco, J. & Wong, K. *Proc. natn. Acad. Sci. U.S.A.* **72**, 1376–1380 (1975).

23. Burt, D. R., Creese, I. & Snyder, S. H. *Molec. Pharmac.* **12**, 800–812 (1976).
24. Leysen, J. E., Gommeren, W. & Laduron, P. M. *Biochem. Pharmac.* **27**, 307–316 (1978).
25. Baudry, M., Martres, M. P. & Schwartz, J. C. in *Catecholamines: Basic and Clinical Frontiers* Vol. 1 (eds Usdin, E., Kopin, I. J. & Barchas, J.) 565–567 (Pergamon, New York, 1978).
26. Martres, M. P., Sokoloff, P. & Schwartz, J. C. in *The Psychopharmacology and Biochemistry of Neurotransmitter Receptors* (eds Olsen, R. N. & Yamamura, H. I.) (Elsevier, Amsterdam, in the press).
27. Trabucchi, M., Longoni, R., Fresia, P. & Spano, P. F. *Life Sci.* **17**, 1551–1556 (1975).
28. Martres, M. P. *et al. Brain Res.* **136**, 319–337 (1977).
29. Di Chiara, G., Porceddu, M. L., Fratta, W. & Gessa, G. L. *Nature* **267**, 270–272 (1977).
30. Lieberman, A. N. *et al. Adv. Neurol.* **24**, 461–473 (1979).
31. Tamminga, C. A., Schaffer, M. H., Smith, R. C. & Davis, J. M. in *Catecholamines: Basic and Clinical Frontiers* Vol. 2 (eds Usdin, E., Kopin, I. J. & Barchas, J.) 1836–1838 (Pergamon, New York, 1978).
32. Parker, R. B. & Waud, D. R. *J. Pharmac. exp. Ther.* **177**, 1–12 (1971).
33. Albe-Fessard, D., Stutinsky, F. & Libouban, S. *Editions du CNRS*, Paris (1971).
34. Moskal, J. R. & Basu, S. *Analyt. Biochem.* **65**, 449–457 (1975).
35. Aures, D., Håkanson, R. & Schauer, A. *Eur. J. Pharmac.* **3**, 217–234 (1968).
36. König, F. R. & Klippel, R. A. *The Rat Brain* (Krieger, Huntington, New York, 1970).
37. Buda, M., Roussel, B., Renaud, B. & Pujol, J. F. *Brain Res.* **93**, 564–569 (1975).
38. Llorens-Cortes, C., Pollard, H. & Schwartz, J. C. *Neurosci. Lett.* **12**, 165–170 (1979).

The enkephalinase inhibitor thiorphan shows antinociceptive activity in mice

B. P. Roques*, M. C. Fournié-Zaluski*, E. Soroça*, J. M. Lecomte†, B. Malfroy‡, C. Llorens‡ & J.-C. Schwartz‡

* Département de Chimie Organique, Université René Descartes, 4, avenue de l'Observatoire, 75006 Paris, France

† Centre de Recherche Janssen-Lebrun, 93 Aubervilliers, France

‡ Unité de Neurobiologie, Centre Paul Broca de l'INSERM, 2ter, rue d'Alésia, 75014 Paris, France

There is both theoretical and therapeutic interest in establishing whether the signals conveyed by the enkephalins are turned off under the action of a specific peptidase which might, in this case, represent a target for a new class of psychoactive agents. Enkephalinase, a dipeptidyl carboxypeptidase cleaving the Gly³-Phe⁴ bond of enkephalins^{1,2} and distinct from angiotensin converting enzyme (ACE)^{3–6}, might be selectively involved in enkephalinergic transmission. It is a membrane-bound enzyme^{1,7,8} whose localization in the vicinity of opiate receptors in the central nervous system is suggested by parallel regional^{1,7,9} and subcellular¹⁰ distributions as well as by the effects of lesions⁹. Such a role is further supported by the ontogenetic development of enkephalinase^{11,12}, its substrate specificity accounting for the increased biological activity of several enkephalin analogues¹³ and its adaptive increase following chronic treatment with morphine^{1,14}. To investigate the functional role of this enzyme further, we have designed a potent and specific enkephalinase inhibitor. We report here that this compound, thiorphan [(DL-3-mercapto-2-benzylpropanoyl)-glycine; patent no. 8008601] protects the enkephalins from the action of enkephalinase *in vitro* in nanomolar concentration and *in vivo* after either intracerebroventricular or systemic administration. In addition, thiorphan itself displays antinociceptive activity which is blocked by naloxone, an antagonist of opiate receptors.

Thiorphan was designed to interact with critical residues of the active site of enkephalinase, as postulated on a hypothetical model (Fig. 1). The model was developed on the assumption that binding of enkephalins to this active site occurs in a manner analogous to that established at the molecular level for carboxypeptidase A (ref. 15). In thiorphan, most of the putative interactions of enkephalins are retained and the binding with the zinc atom at the catalytic site is strongly increased by the introduction of a thiol group in the proper position^{16,17}.

Evidence for the validity of the hypothetical model of the enkephalinase active site was suggested by the high inhibitory potency of thiorphan (Fig. 2). The inhibition of enkephalinase activity from striatal membranes is of a competitive type and the mean K_i value obtained in various experiments similar to those described in Fig. 2 was 4.7 ± 1.2 nM.

Furthermore, thiorphan displays a good selectivity towards enkephalinase among the various enkephalin-hydrolysing

activities already described in brain tissues. Regarding ACE activity in mouse striatum¹⁸ which is also able to cleave the Gly³-Phe⁴ peptide bond of enkephalins¹⁹, the K_i of thiorphan was 150 ± 30 nM (not shown) which corresponds to a 30-fold lower potency than towards enkephalinase. It is interesting that captopril (SQ 14,225), an inhibitor of ACE, displays an inverse specificity pattern, being much more potent on this enzyme ($K_i = 7$ nM) than on enkephalinase ($K_i = 10 \mu\text{M}$)^{4,16}. These observations further demonstrate that enkephalinase and ACE, in spite of their analogies³, are different enzymes^{4,14}. In addition, the formation of either ³H-Tyr or ³H-Tyr-Gly, as detected by TLC, following incubation of striatal membranes in the presence of 20 nM ³H-Leu⁵-enkephalin and in the absence of puromycin (which inhibits the formation of both compounds) was not affected by thiorphan in concentrations as high as 10 μM . This indicates that thiorphan does not significantly inhibit the aminopeptidase^{20–26} or the putative endopeptidase⁶ activities responsible for *in vitro* cleavage of the enkephalin molecules at peptide bonds other than Gly³-Phe⁴.

The ability of thiorphan to inhibit enkephalinase activity *in vivo* was evaluated by determining the effect of the drug on the analgesia elicited by (D-Ala²-Met⁵)-enkephalin administered intracerebroventricularly (i.c.v.) in low dosage. The D-Ala² substitution in this compound preserves opiate receptor affinity and bestows resistance to aminopeptidases²⁷, but does not markedly impair recognition by enkephalinase as long as the carboxyl of Met⁵ is not amidified¹³.

A short-lasting, nonsignificant analgesia was recorded on the tail-withdrawal test²⁸ when (D-Ala²-Met⁵)-enkephalin was administered (20 μg , i.c.v.) alone (Fig. 3 and Table 1). Co-administration of thiorphan, either i.c.v. or systemically, resulted in a strong potentiation of the analgesic effect of the opioid pentapeptide which was nearly maximal in all animals over ~90 min and still significant ~2 h after intravenous (i.v.) administration and 4 h after i.c.v. administration. In contrast, the effect elicited by (D-Ala²-Met⁵)-enkephalinamide, an analogue poorly recognized by enkephalinase¹³, was not significantly modified by thiorphan. Interestingly, neither thiorphan alone (30 μg or 60 μg , i.c.v.) nor the opiate receptor antagonist naloxone (1 mg per kg, subcutaneously (s.c.))

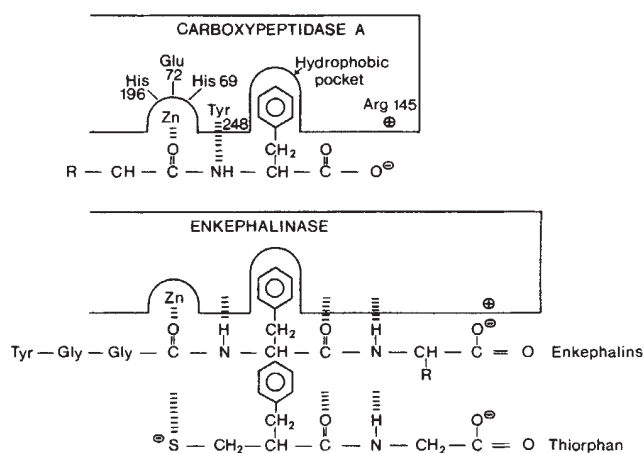


Fig. 1 Hypothetical model of the active site of enkephalinase and of the binding of enkephalins and thiorphan. The model was developed by analogy with the active site of carboxypeptidase A (upper part) as proposed by Quiocho and Lipscomb¹⁵. Enkephalinase is likely to contain a zinc atom³ that, as in carboxypeptidase A, would bind the carbonyl group of the peptide bond to be cleaved. Evidence for the presence of a 'dead-end hydrophobic pocket' analogous to that of carboxypeptidase A¹⁵ but which seems absent in ACE¹⁶ is given by the recognition of di- and tripeptides only when containing an aromatic amino acid³⁷. Interaction of the two C-terminal peptide bonds of enkephalins with the active site is suggested by the diminished recognition of analogues in which they are modified³⁷. The free carboxyl would interact with a positively charged residue (corresponding to Arg 145 in carboxypeptidase A) as shown by the diminished recognition of enkephalin analogues in which it is replaced by an alcohol group, amidified, esterified or localized in a proline ring^{13,14}.

Fig. 2 Inhibition of enkephalinase activity from mouse striatum by thiorphan. A washed particulate fraction from mouse striatum was prepared according to Malfroy *et al.*¹. Protein concentration in the assay mixture was 40 µg per 100 µl but similar results were obtained for concentrations varying over 1–200 µg per 100 µl. After a 15-min preincubation at 25 °C, incubations were started by addition of a mixture containing 20 nM (³H-Leu⁵)-enkephalin previously purified by Porapak column chromatography²⁴, 0.1 mM puromycin, an aminopeptidase inhibitor^{9,25}, 1 µM captopril, an ACE inhibitor¹⁶, 0.1 mM 2-mercaptoethanol and unlabelled (Leu⁵)-enkephalin in the indicated concentrations (all final concentrations in 50 mM Tris-HCl buffer, pH 7.4). Incubations (15 min at 25 °C) were stopped by addition of HCl (0.07 M, final molarity). Blanks were obtained by adding HCl before the beginning of incubation and represented ~20% of the uninhibited enzyme activity. ³H-Labelled metabolites plus unlabelled carriers were separated by TLC on silica gel plastic sheets (Merck, Darmstadt) using the following system: BuOH, AcOH, AcOEt, H₂O (1:1:2:1). Spots, localized after a ninhydrin spray, had the following *R_F* values: Tyr 0.55, Tyr-Gly 0.47, Tyr-Gly-Gly 0.39, (Leu⁵)-enkephalin 0.81. The Tyr-Gly-Gly spots were cut out and their radioactivity estimated by liquid scintillation counting. Left: Effect of thiorphan in increasing concentrations on the hydrolysis of (³H-Leu⁵)-enkephalin (20 nM) by mouse striatal enkephalinase. Means ± s.e.m. of triplicate determinations in the same experiment. This experiment was repeated 12 times leading to an IC₅₀ value of 4.7 ± 1.2 nM. Right: Dixon plots of enkephalinase inhibition by thiorphan. Each point represents the mean of duplicate determinations. *K_i* value obtained in this experiment was 3 nM and was confirmed in another similar experiment.

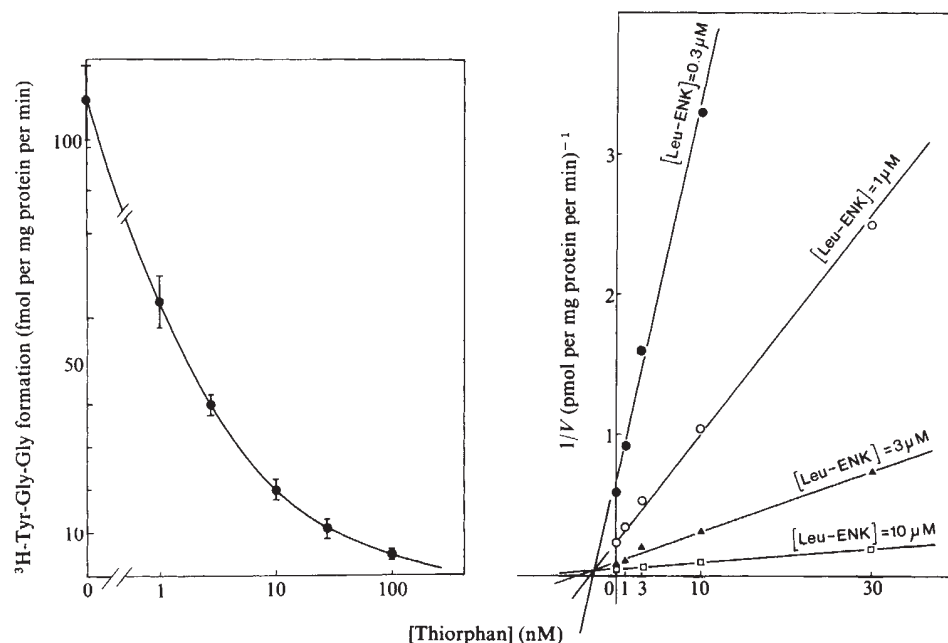


Table 1 Activity of thiorphan on responses to nociceptive stimuli in mice

Tests	Treatments	Latency (s)	<i>P</i>
Tail withdrawal (48 °C)	Controls	3.1 ± 0.4 (15)	
	Thiorphan (30 µg, i.c.v.)	4.2 ± 1.1 (6)	
	Thiorphan (60 µg, i.c.v.)	2.7 ± 0.4 (9)	
	Naloxone (1 mg per kg, i.v.)	3.2 ± 0.3 (9)	
	(D-Ala ² -Met ⁵ -enkephalin (20 µg, i.c.v.))	3.4 ± 0.2 (14)	
	(D-Ala ² -Met ⁵ -enkephalin (20 µg, i.c.v.) + thiorphan (30 µg, i.c.v.))	9.6 ± 0.9 (14)	<0.001
	(D-Ala ² -Met ⁵ -enkephalin (20 µg, i.c.v.) + thiorphan (100 mg per kg, i.v.))	7.6 ± 0.8 (14)	<0.001
	(D-Ala ² -Met ⁵ -enkephalinamide (2.5 µg, i.c.v.))	3.2 ± 0.4 (6)	
	(D-Ala ² -Met ⁵ -enkephalinamide (2.5 µg, i.c.v.) + thiorphan (30 µg, i.c.v.))	3.3 ± 0.6 (6)	
	(D-Ala ² -Met ⁵ -enkephalinamide (15 µg, i.c.v.))	12.5 ± 1.1 (6)	<0.001
	(D-Ala ² -Met ⁵ -enkephalinamide (15 µg, i.c.v.) + thiorphan (30 µg, i.c.v.))	13.8 ± 1.2 (6)	<0.001*
Hotplate jump (55 °C)	Controls (solvent, i.c.v.)	70 ± 6 (14)	
	Thiorphan (30 µg, i.c.v.)	121 ± 8 (10)	<0.001
	Thiorphan (60 µg, i.c.v.)	134 ± 17 (14)	<0.01
	Controls (solvent, i.v.)	63 ± 11 (20)	
	Thiorphan (100 mg per kg, i.v.)	129 ± 22 (10)	<0.02
	Thiorphan (100 mg per kg, i.v.) + naloxone (1 mg per kg, s.c.)	68 ± 16 (10)	<0.05†
Hotplate jump (50 °C)	Naloxone (1 mg per kg, s.c.)	43 ± 8 (10)	<0.05
	Controls	186 ± 11 (20)	
	Thiorphan (60 µg, i.c.v.)	245 ± 14 (20)	<0.05
	Thiorphan (60 µg, i.c.v.) + naloxone (1 mg per kg, s.c.)	129 ± 13 (20)	<0.001†

The tail withdrawal test was performed as described in Fig. 3 legend. The latency values obtained for each animal during the 240-min experimental period (13 determinations) were averaged. The hotplate test^{34,35} was carried out at either 55 °C or 50 °C and the jump latency recorded 15 min after drug administration, without 'cutoff time'. Naloxone was administered at the same time as thiorphan. Results are the means ± s.e.m. of the number of experiments indicated in parentheses.

* Nonsignificant compared with (D-Ala²-Met⁵)-enkephalin alone.

† Compared with thiorphan alone.

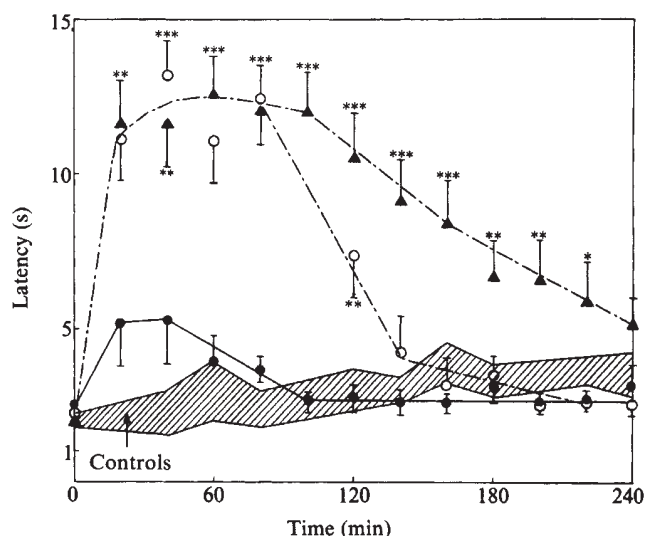


Fig. 3 Effects of thiorphan on the analgesic activity of (D-Ala²-Met⁵)-enkephalin in the mouse tail withdrawal test. Swiss mice (20–22 g, Charles River) were tested at 20-min intervals over 4 h. The time (in seconds) between tail immersion into a water bath maintained at 48 °C and withdrawal was recorded as the response latency²⁸. The 'cutoff time' was 15 s and the value at zero time represents the average of three determinations carried out before drug injection. (D-Ala²-Met⁵)-enkephalin (20 µg) was administered i.c.v. under a volume of 5 µl (3% ethanolic solution) either alone or together with thiorphan (30 µg). In the third group of animals thiorphan was administered i.v. (100 mg per kg) 15 min before (D-Ala²-Met⁵)-enkephalin (20 µg, i.c.v.). Controls received an i.c.v. administration of solvent alone. Each point represents the mean ± s.e.m. of data from 14 or 15 experiments. The shaded area corresponds to means ± s.e.m. of controls. ▲, (D-Ala²-Met⁵)-enkephalin + thiorphan (i.c.v.); ○, (D-Ala²-Met⁵)-enkephalin + thiorphan (i.v.); ●, (D-Ala²-Met⁵)-enkephalin alone. * *P* < 0.05; ** *P* < 0.01; *** *P* < 0.001.

modified the responsiveness of mice in the tail-withdrawal test, at any time after administration (Table 1).

Because the involvement of endogenous opioid peptides in the response to various nociceptive stimuli, as revealed by the hyperalgesic effect of naloxone, seems to vary according to the type of nociceptive stimulus used²⁹⁻³³, we have determined the effect of thiorphan in the hotplate jump test^{34,35}. On this test, performed at either 55 °C or 50 °C, thiorphan administered either i.c.v. or i.v. displayed a clear antinociceptive activity which was probably mediated by opiate receptors as it was completely prevented by administration of naloxone in moderate dosage (Table 1). This opiate antagonist, when administered alone, significantly decreased the latency of the jump response as previously shown^{29,32}.

The effect of thiorphan cannot be mediated by inhibition of aminopeptidase and is unlikely to involve ACE inhibition because its antinociceptive activity was observed at dosages approximately 30 times lower than those required for captopril³⁶, whereas the latter compound is approximately 20 times more potent than thiorphan as an ACE inhibitor. Hence, the present observations strongly support the hypothesis that enkephalinase is selectively responsible for the metabolism of endogenous enkephalins. The reasons for which thiorphan elicits analgesia (and naloxone hyperalgesia) in one test but not in others are unclear but might be that different noxious stimuli elicit release of endogenous opioid peptides to a variable extent. Thus, thiorphan might constitute a useful tool for investigating the involvement of enkephalins in a variety of behavioural situations and might also represent the prototype of a new class of pharmacological agent.

We thank J. L. Coignet, M. Schwob and M. De Munck for technical assistance. This work was supported by grants from the Délégation Générale à la Recherche Scientifique et Technique, the Centre National de la Recherche Scientifique (PIRMED), the Institut National de la Santé et de la Recherche Médicale, and the Fondation pour la Recherche Médicale.

Received 6 June; accepted 17 September 1980.

- Malfroy, B., Swerts, J. P., Guyon, A., Roques, B. P. & Schwartz, J.-C. *Nature* **276**, 523-526 (1978).
- Guyon, A. *et al. Life Sci.* **25**, 1605-1612 (1979).
- Swerts, J. P., Perdriset, R., Malfroy, B. & Schwartz, J.-C. *Eur. J. Pharmac.* **53**, 209-210 (1979).
- Swerts, J. P., Perdriset, R., Patey, G., De La Baume, S. & Schwartz, J.-C. *Eur. J. Pharmac.* **57**, 279-281 (1979).
- Sullivan, S., Akil, H., Blacker, D. & Barchas, J. D. in *Endogenous and Exogenous Opiate Agonists and Antagonists* (ed. Way, E. L.) 357-360 (Pergamon, New York, 1980).
- Gorenstein, C. & Snyder, S. H. *Life Sci.* **25**, 2065-2070 (1979).
- Sullivan, S., Akil, H. & Barchas, J. D. *Commun. Psychopharmac.* **2**, 525-531 (1979).
- Vogel, Z. & Alstein, M. in *Endogenous and Exogenous Opiate Agonists and Antagonists* (ed. Way, E. L.) 353-356 (Pergamon, New York, 1980).
- Malfroy, B., Swerts, J. P., Llorens, C. & Schwartz, J.-C. *Neurosci. Lett.* **11**, 329-334 (1979).
- De La Baume, S., Patey, G. & Schwartz, J.-C. *Neuroscience* (in the press).
- De La Baume, S., Patey, G., Gros, C. & Schwartz, J.-C. in *Endogenous and Exogenous Opiate Agonists and Antagonists* (ed. Way, E. L.) 179-182 (Pergamon, New York, 1980).
- Patey, G., De La Baume, S., Gros, C. & Schwartz, J.-C. *Life Sci.* **27**, 245-252 (1980).
- Fournie-Zaluski, M. C. *et al. Biochem. biophys. Res. Commun.* **91**, 130-135 (1979).
- Schwartz, J.-C. *et al. Adv. Biochem. Psychopharmac.* **22**, 219-235 (1980).
- Quirocho, F. A. & Lipscomb, W. N. *Adv. Protein Chem.* **15**, 1-48 (1971).
- Ondetti, M. A., Rubin, B. & Cushman, D. W. *Science* **196**, 441-444 (1977).
- Ondetti, M. A. *et al. Biochemistry* **18**, 1427-1430 (1979).
- Yang, H. Y. T. & Neff, N. H. *J. Neurochem.* **19**, 2443-2450 (1972).
- Erdos, E. G., Johnson, A. R. & Boyden, N. T. *Biochem. Pharmac.* **27**, 843-848 (1978).
- Hambrook, J. M., Morgan, B. A., Rance, M. J. & Smith, C. F. *Nature* **262**, 782-783 (1976).
- Lane, A. C., Rance, M. J. & Walter, D. S. *Nature* **269**, 75-76 (1977).
- Marks, N., Grynbaum, A. & Neidle, A. *Biochem. biophys. Res. Commun.* **74**, 1552-1559 (1977).
- Meek, J. L., Yang, H. Y. T. & Costa, E. *Neuropharmacology* **16**, 151-154 (1977).
- Vogel, Z. & Alstein, M. *FEBS Lett.* **80**, 332-336 (1977).
- Knight, M. & Klee, W. A. *J. Biol. Chem.* **253**, 3843-3847 (1978).
- Guyon, A. *et al. Biochem. biophys. Res. Commun.* **88**, 919-926 (1979).
- Pert, C. B., Pert, A., Chang, J. K. & Fong, B. T. W. *Science* **194**, 331-332 (1976).
- Ben-Bassat, J., Peretz, E. & Sulman, F. G. *Archs int. Pharmacodyn. Ther.* **122**, 434-477 (1959).
- Jacob, J., Tremblay, E. C. & Colombel, M. C. *Psychopharmacologia* **37**, 217-223 (1974).
- Goldstein, A., Pryor, G. T., Otis, L. S. & Larson, F. *Life Sci.* **18**, 599-604 (1976).
- Sewell, R. D. E. & Spencer, P. S. *J. Neuropharmacology* **15**, 683-688 (1976).
- Frederickson, R. C. A., Bergers, V. & Edwards, J. D. *Science* **198**, 756-757 (1977).
- Wolf, C. J. *Brain Res.* **190**, 578-583 (1980).
- Eddy, N. B., May, E. L. & Mosettig, E. *J. Org. Chem.* **17**, 321-326 (1952).
- Jacob, J. & Blozovski, M. *Archs int. Pharmacodyn. Ther.* **133**, 296-309 (1961).
- Stine, S. M., Yang, H. Y. T. & Costa, E. *Brain Res.* **188**, 295-299 (1980).
- Llorens, C. *et al. Biochem. biophys. Res. Commun.* (in the press).

Herpes simplex virus DNA sequences in the CNS of latently infected mice

Carlos V. Cabrera*, Charles Wohlenberg, Harry Openshaw†, Manuel Rey-Mendez‡, Alvaro Puga & Abner Louis Notkins

Laboratory of Oral Medicine, National Institute of Dental Research, National Institutes of Health, Bethesda, Maryland 20205

It has been amply documented that herpes simplex virus (HSV) persists in sensory ganglia of the peripheral nervous system (PNS)¹⁻⁵. In contrast, HSV latency in the central nervous system (CNS) has not been well characterized. Corneal inoculation of virus results in a productive viral infection in the CNS during the first week after inoculation⁶, indicating that the virus can progress from the PNS to the CNS. During latency, HSV has been found by co-cultivation of CNS tissue in only a very small fraction of inoculated mice^{7,8}. We have used here molecular hybridization techniques to analyse the fate of viruses that reach the CNS by anatomical pathways. We show that 6 days after corneal inoculation of HSV-1 a productive viral infection was present in brain tissue as well as in peripheral ganglia in at least 90% of the inoculated mice. The mortality during this acute phase was only 2%. In the survivors, latent HSV could be recovered by explantation from 95% of the trigeminal ganglia, but only 5% of the brain tissue explants of the same mice yielded infectious virus. However, HSV DNA sequences were detected in the brains of 30% of mice which harboured latent HSV in their trigeminal ganglia. These results suggest that viruses that progress from the PNS into the CNS are not eliminated, but are capable of establishing a latent infection in the CNS that cannot be reactivated by explantation techniques.

Six- to eight-week-old female BALB/cJ mice were inoculated by corneal scarification with approximately 5×10^6 plaque-forming units (PFU) of the F strain of type 1 HSV (American Type Culture Collection). The virus was grown in primary rabbit kidney cells and concentrated by ultracentrifugation (stock pool of 10^8 PFU ml⁻¹). For infectivity assays, homogenates or minced explants of individual organs were monitored for ability to induce cytopathic effects in monolayers of primary rabbit kidney cells³⁻⁵. For hybridization experiments, DNA was extracted from individual brain stems and hemispheres. Hybridization was to a ³²P-labelled HSV DNA probe prepared by nick translation⁹, purified in alkaline Sephadex G-200 (ref. 10) and free of fold-back DNA⁴. The conditions for the DNA-DNA hybridization reactions were as described previously^{4,11,12}.

Table 1 shows that infectious HSV could be detected in homogenates of trigeminal ganglia from all of the mice killed 6 days after corneal inoculation of HSV (acute stage). At this time, infectious HSV was also detected in the brain stems of 18 out of 20 mice, and in the brain hemispheres of 6 out of 20 mice. DNA-DNA hybridization confirmed and extended these findings. Figure 1a shows the hybridization kinetics of a trace of ³²P-HSV DNA with DNA from individual brain stems and hemispheres taken from mice 6 days after inoculation. All 10 DNA preparations increased the rate of reassociation of the tracer over the self-reassociation rate in the presence of uninfected mouse brain DNA. However, the extent of the rate increase was not the same for the different preparations tested. The slopes of the curves show that the brain stems from acutely infected mice contained more viral DNA than the hemispheres from the same mice. The range of viral equivalents per 100 cells was 1-5 for the brain hemispheres and 22-178 for the brain stems (Table 2). In each animal the brain stem had more viral

Present addresses: * Division of Biology, California Institute of Technology, Pasadena, California 91125; † Dept of Neurology, UC Davis Medical Center, Sacramento, California 95817; ‡ Depto de Bioquímica, facultad de Ciencias Biológicas, Universidad de Santiago, Santiago, Spain.

Table 1 Infectivity assay of mouse ganglia and brains

Time after inoculation	Trigeminal ganglia	Brain stem	Brain hemispheres
6 days			
Homogenates*	20/20	18/20	6/20
8 weeks			
Homogenates*	0/20	0/20	0/20
Explants†	38/40	0/20	1/20

Results are expressed as the ratio of positive isolates from individual mice to the total number of mice tested.

* The appropriate tissue was homogenized in glass Ten-Broek tissue grinders. After freezing and thawing to disrupt viable cells, the cell debris was pelleted by low-speed centrifugation and the supernatant fluid was cultured on primary rabbit kidney cells. Cultures were followed for viral-induced cytopathology (CPE) for 5 days.

† Ganglia were explanted on monolayers of primary rabbit kidney cells in Eagle's minimal essential media supplemented with 3% fetal bovine serum (MEM). CPE appeared in the cell monolayer by 8 days in over 95% of the explant cultures that yielded HSV. For CNS explants, the appropriate tissues were minced and placed in MEM without a cell monolayer. The media was changed on alternate days and cultured for HSV. After 6–8 days, the explant tissue was homogenized and assayed for infectious virus*. As shown previously, this technique can detect *in vitro* reactivation in 100% of latently infected ganglia after only 48 h of culture⁵. In contrast, CPE does not usually appear on the indicator cell monolayer before 5 days in the standard explantation and co-cultivation technique⁷.

sequences than the brain hemispheres, ranging from 8-fold for animal no. 4 to 60-fold for animal no. 3.

Mortality in the inoculated animals was only 2%, despite the productive viral infection in the CNS. Furthermore, 8 weeks after inoculation, trigeminal ganglionic homogenates were negative, but 38 of 40 ganglionic explants yielded infectious HSV by *in vitro* reactivation. Although these conditions essentially define the latent stage of the ganglionic infection, they did not apply to the CNS tissue, as HSV could be recovered by explantation in only 1 of 20 mice (Table 1). This low recovery rate from CNS explants could be explained in two ways: either the virus is eliminated from the CNS during the acute infection or the viral genome is present in the CNS in a form that cannot be readily reactivated by explantation techniques. To investigate these alternatives, a total of 20 mice in three separate inoculation schedules were killed 8 weeks after inoculation with HSV. DNA was extracted from the brain stems and hemispheres and each individual preparation was hybridized to a ³²P-HSV DNA probe.

The highest possible mass ratios of brain DNA to probe were used to maximize the detection level. However, due to the limited amount of DNA available from single organs, the sensitivity level was never higher than 1 viral equivalent per 1,500 cells, and usually it was in the range of 1 viral equivalent per 300–800 cells. Notwithstanding, DNA from 3 out of 20 hemispheres and 4 out of 20 stems increased significantly the reassociation rate of the probe over the control (Fig. 1b). An average number of copies per cell was calculated for the seven positive organs using the rates determined experimentally and the known parameters of each hybridization reaction. The three positive hemispheres had 0.4–9 copies per 100 cells; the four positive stems had 0.5–6 copies per 100 cells. Unlike the acute stage, latently infected brain stems and hemispheres had comparable amounts of viral DNA, although the number of sequences in either tissue varied over a 10–20-fold range (Table 2; only positive cases shown). These calculations are based on the assumption that the entire viral genome is present, although the maximum extent of hybridization never exceeded 40% (H-8) and more often was 12–20% (S-9; S-11). This assumption is only valid to determine an average number of copies for the DNA sequences that hybridize. Only one animal (no. 7) contained viral sequences in both brain stem and hemispheres; all other positive animals had viral DNA in either tissue, but not both.

Latent HSV has previously been detected by co-cultivation in 9% of mouse brains following experimentally induced viraemia⁷ and in 9% of mouse spinal cords after rear-footpad inoculation⁸. In those reports mortality was high, unlike the present study. However, in our experiments, 5% of the brain hemispheres were shown to contain virus, a number not very different from those reported previously. Clearly, explantation results in a large underestimation of the actual number of animals that harbour viral DNA sequences in the CNS. As shown, a total of 6 out of 20 mice (30%) have HSV DNA in the brain stem and/or hemispheres, a fraction six times higher than that found by explantation. Inoculation of a viral dose sufficiently large to increase the mortality rate may increase the fraction of surviving animals with detectable viral DNA sequences in the CNS.

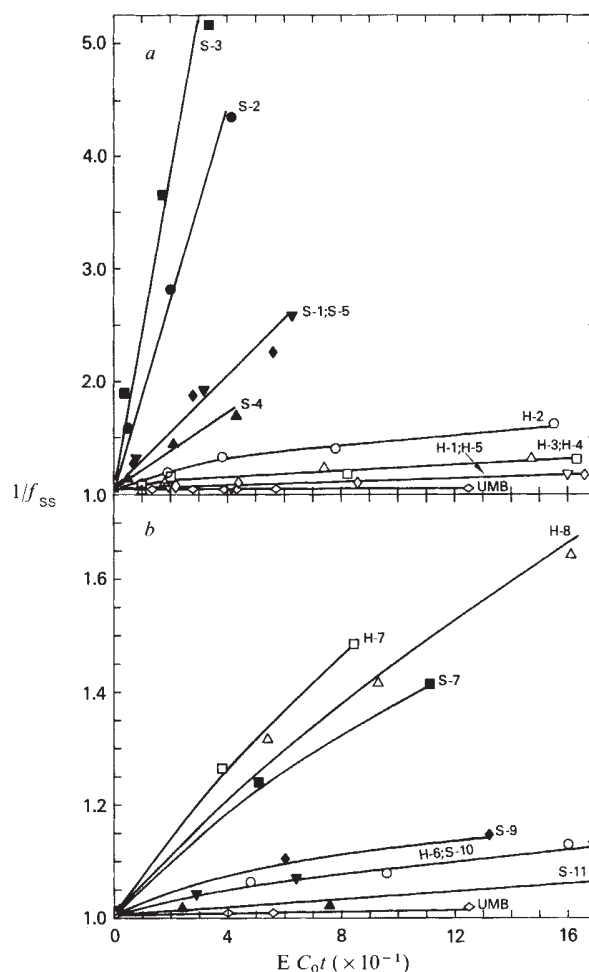


Fig. 1 Hybridization kinetics of ³²P-HSV DNA (1.2×10^8 c.p.m. per μ g) in the presence of DNA from brain stems (S) or brain hemispheres (H) of HSV-infected mice. Brain tissues were homogenized in 2.2 M sodium trichloroacetate (NaTCA), pH 7.4, 10 mM EDTA, followed by phenol-chloroform extraction and ethanol precipitation. NaTCA inhibits DNase activity present in large quantities in nervous tissue. DNA preparations were treated with RNase, proteinase K, re-extracted and sonicated to an average piece size of 7S in alkaline sucrose. DNA yield was 300 ± 50 μ g for brain hemispheres and 40 ± 15 μ g for brain stems. **a**, 6 days post-inoculation; the reaction mixtures contained 0.8–1.0 ng ml⁻¹ ³²P tracer and 11–12 mg ml⁻¹ brain hemisphere DNA or 5–6 mg ml⁻¹ brain stem DNA. **b**, 8 weeks post-inoculation; the reaction mixtures contained 1 ng ml⁻¹ ³²P tracer and 15 mg ml⁻¹ hemisphere DNA or 13–20 mg ml⁻¹ stem DNA. A control was run in the presence of uninfected mouse brain DNA (UMB). Mixtures were adjusted to 0.48 or 1.0 M phosphate buffer, pH 6.8, and incubated in siliconized glass capillaries at 70–72 °C. The inverse of the fraction of ³²P-DNA remaining as single strands ($1/f_{ss}$) is plotted as a function of the equivalent C_0t ($E C_0t$), corrected to 0.18 M Na⁺ (ref. 12).

The detection level of the hybridization reactions is limited by the scarcity of DNA from single CNS tissues and by the uncertainty as to which areas of the brain are most likely to be positive for viral sequences. The hybridization rates used to calculate the number of viral copies have been obtained from a least-squares regression fit of three experimental points to a straight line, but in many cases a fit to a two-component curve may also be possible (as shown in Fig. 1). A fit to a straight line provides only valid information for comparisons to be drawn between the various tissues. Thus, the results would not warrant the conclusion that all of the viral genome is present in the CNS during the latent stage. In fact, the sequences detected may correspond to an unbalanced representation of portions of the genome, such as is found in defective particles. The presence of defective particles may explain the absence of significant reactivation by explantation.

HSV reaches sensory ganglia in the PNS by intra-axonal retrograde transport¹³⁻¹⁵. The same anatomical pathway is the most likely route for virus to reach the CNS, as the bipolar axon of trigeminal neurones continues past the ganglion and synapses into the brain stem. A productive viral infection in the brain stem and hemispheres can then occur concomitantly with the acute stage of the ganglionic infection, setting the stage for subsequent latency in the CNS.

The finding of latent viral DNA in the brains of experimental animals suggests a possible role for HSV in neurological disorders of unknown aetiology. If the viral genome or portions of it were shown to be present in human brains, in the absence of complete viral reactivation certain genes could be expressed

under the influence of various stimuli. The resulting viral proteins might interfere with nerve cell function by altering cell membranes, setting up abnormal storage material or provoking a local immune response.

We thank Dr Edouard Cantin for helpful discussions, Drs Patrick McClintock and Martin Haspel for criticisms, Mrs Jeannette Murphy and Edith Rian for secretarial help, and Dr D. E. Kohne for communicating methodology before its publication.

Received 3 July; accepted 17 September 1980.

1. Stevens, J. G. & Cook, M. L. *Science* **173**, 843 (1971).
2. Price, R. W., Katz, B. J. & Notkins, A. L. *Nature* **257**, 686 (1975).
3. Walz, M. A., Price, R. W. & Notkins, A. L. *Science* **184**, 1185 (1974).
4. Puga, A., Rosenthal, J. D., Openshaw, H. & Notkins, A. L. *Virology* **89**, 102 (1978).
5. Openshaw, H., Asher, L. S. V., Wohlenberg, C., Sekizawa, T. & Notkins, A. L. *J. gen. Virol.* **44**, 205 (1979).
6. Baringer, J. R. & Griffith, J. F. *J. Neuropath. exp. Neurol.* **29**, 89 (1970).
7. Knotts, E. P., Cook, M. L. & Stevens, J. G. *J. exp. Med.* **138**, 740 (1973).
8. Cook, M. L. & Stevens, J. G. *J. gen. Virol.* **31**, 75 (1976).
9. Maniatis, T., Jeffrey, A. & Kleid, D. G. *Proc. natn. Acad. Sci. U.S.A.* **72**, 1184 (1975).
10. Britten, R. J., Graham, D. E. & Neufeld, B. B. *Meth. Enzym.* **29**, 363 (1974).
11. Kohne, D. E. *Biophys. J.* **8**, 1104 (1968).
12. Britten, R. J. & Smith, J. *Yb. Carnegie Instn Wash.* **68**, 378 (1970).
13. Hill, T. J., Field, H. J. & Roome, A. P. C. *J. gen. Virol.* **15**, 253 (1972).
14. Cook, M. L. & Stevens, J. G. *Infect. Immun.* **7**, 272 (1973).
15. Baringer, J. R. & Swoveland, P. *Lab. Invest.* **30**, 230 (1974).

Immunoglobulin C_μ RNA in T lymphoma cells is not translated

Ian D. Walker & Alan W. Harris

The Walter and Eliza Hall Institute of Medical Research,
Post Office, Royal Melbourne Hospital, Victoria 3050, Australia

Table 2 Viral DNA sequences in brain stems and hemispheres of HSV-infected mice

Time after inoculation	CNS tissue and animal no.	Rate increase	DNA copies per 100 cells
6 days	Brain hemispheres		
	1	5 ± 1	1 ± 0.3
	2	19 ± 2	5 ± 0.6
	3	9 ± 1	3 ± 0.4
	4	9 ± 1	3 ± 0.5
	5	5 ± 2	1 ± 0.5
	Brain stem		
	1	61 ± 5	38 ± 3
	2	138 ± 11	120 ± 10
	3	154 ± 21	178 ± 24
	4	27 ± 3	22 ± 3
	5	42 ± 9	28 ± 6
8 weeks	Brain hemispheres		
	6	2 ± 0.3	0.4 ± 0.2
	7	55 ± 7	9 ± 1
	8	4 ± 0.5	3 ± 0.4
	Brain stem		
	7	38 ± 6	6 ± 1
	9	29 ± 9	2 ± 0.5
	10	8 ± 1	1 ± 0.2
	11	5 ± 1	0.5 ± 0.1
	R-15	15 ± 2	—
	R-24	24 ± 7	—

Rate increase factors and copies per cell were determined as described⁴, using $2.46 \pm 0.17 \text{ l mol}^{-1} \text{ s}^{-1}$ as the rate constant for the reassociation of nick-translated HSV DNA in the presence of uninfected mouse brain DNA (an average of five experiments). The rate constants of the hybridization reactions were determined from the least-squares fit of the curves in Fig. 1. A total of 20 brain stems and their corresponding hemispheres were tested at 8 weeks post-inoculation; only organs that gave rate increase ≥ 2 were considered positive and are reported here. The mass ratios of brain DNA to ³²P-HSV DNA ranged from $5 \times 10^6:1$ to $1 \times 10^8:1$. Results are indicated $\pm$ s.e. R-15 and R-24 are reconstitution experiments containing 15-fold and 24-fold excess of unlabelled HSV DNA over ³²P tracer, respectively; the rate increases in these two reactions agree with those expected from the reconstitution conditions.

It is widely believed that immunoglobulin genes might encode at least part of the receptor for antigen on the T lymphocyte. Evidence supporting this comes from the effects of anti-immunoglobulin idiotype antibodies on cellular immune networks¹⁻⁴ and from the presence of idiotypes on immunologically active factors from T cells⁵⁻⁷. Detailed molecular characterization of the receptor, however, has been seriously hampered by the lack of a suitable cellular source from which it might be isolated. The recent demonstration by Kemp *et al.*^{8,9} that thymocytes and certain cultured lines of mouse T lymphoma cells contain polyadenylated RNA molecules encoded by the immunoglobulin C_μ gene (C_μ RNA) prompted us to identify the corresponding protein molecules in those cells. As the haploid mouse genome contains a single C_μ gene¹⁰, any polypeptide encoded by this gene should react with at least some of the antibodies present in rabbit anti-mouse IgM antiserum. In this letter we report that a number of T lymphoma lines, regardless of whether they contain C_μ RNA, synthesize no detectable μ polypeptides.

The cloned T lymphoma lines chosen for study were STRij-4-2.2 (ST4), STRij-1.3 (ST1), WEHI-22.1 (W22), EL-4.1 (EL4) (described in ref. 8) and WEHI-112.1 (W112): they contained 33, <1.5, 3, <1.5 and 30–50 molecules of C_μ RNA per cell, respectively. In addition, two B lymphoma cell lines, WEHI-231 (W231) (μ, κ) and WEHI-279 (W279) (μ, κ), a mastocytoma line P815 (respectively, 95, 195 and <0.3 C_μ RNA molecules per cell⁸) and an Abelson pre-B lymphoma cell line 18-81 (ref. 11) were used as positive or negative controls in the search for μ polypeptide chains.

The presence of C_μ RNA molecules in ST4 and W112 was repeatedly confirmed during experiments described here, indicating that C_μ gene transcription occurs continuously in these cell lines.

A 'sandwich' radioimmunoassay (RIA) technique using ¹²⁵I-labelled antibodies was used to test cell extracts for μ polypeptide content and Fig. 1a shows the results of an experiment in which extracts of W231, P815 and ST4 were examined. No binding whatever was apparent with ST4 or P815, whereas in

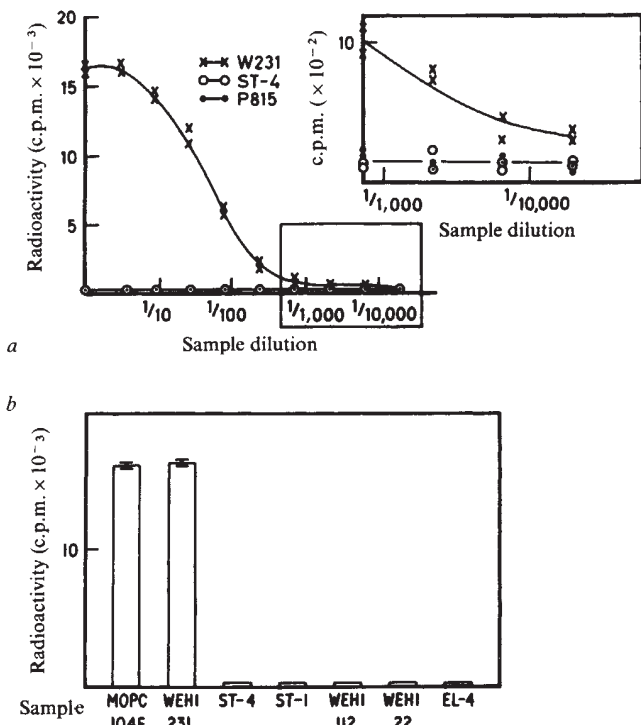


Fig. 1 Detection of μ polypeptides in detergent lysates of lymphoma cells. Cells (10^8 ml^{-1}) were solubilized in Triton X-100 buffer (0.01 M Tris-HCl, 0.15 M NaCl, 0.5% Triton X-100, pH 7.5) containing 1% (w/v) bovine serum albumin (BSA), by incubation at 0°C for 20 min. Nuclei were removed by centrifugation at 10,000 r.p.m. for 30 min. Supernatants were tested by solid-phase RIA¹⁶. Anti- μ antiserum, obtained from rabbits immunized with purified myeloma HPC-76 IgM (μ, κ), was affinity-purified on myeloma MOPC 104E IgM (μ, λ)-Sephacrose adsorbent. Microtitre plates were coated with purified anti- μ Ig ($20 \mu\text{g ml}^{-1}$) and then incubated with dilutions of cell extracts or of purified 104E IgM. After washing, anti- μ ^{125}I -Ig was added ($10 \mu\text{Ci } \mu\text{g}^{-1}$, 20,000 c.p.m. per well), and incubation carried out overnight. Plates were washed and individual wells counted in a Packard γ counter. *a*, Titration curves for extracts of W231, P815 and ST4; *b*, binding data for detergent extracts of various cell lines, each tested at a concentration of $10^8 \text{ cells ml}^{-1}$ compared to MOPC 104E at $10 \mu\text{g ml}^{-1}$.

W231, activity (IgM) could be detected even after a 10,000-fold dilution. Using purified myeloma MOPC 104E IgM as a reference standard, the amount of IgM in W231 was estimated at 3×10^5 molecules per cell. Thus the number of μ molecules in P815 or ST4 cannot exceed 30 per cell. Other C_μ RNA-containing T lymphoma lines (W112 and W22) were tested by sandwich RIA and again no μ polypeptides were detected (Fig. 1*b*).

We considered the following possibilities to account for our failure to detect μ polypeptides in cells producing C_μ RNA:

- (1) The anti- μ antibodies bind only to carbohydrate determinants present on IgM but not on T lymphoma cell μ polypeptides.
- (2) A specific protease in T lymphoma cells degrades μ polypeptides.
- (3) The anti- μ antibodies cannot bind μ chains that are not associated with light chains.
- (4) μ polypeptides in T lymphoma cells are associated with other proteins which render antigens, which would otherwise bind anti- μ antibodies, inaccessible.
- (5) A fragment of μ chain containing only one or a few antigenic sites is synthesized by T lymphoma cells.

Possibility (1) was eliminated by demonstrating that oxidation of HPC-76 myeloma IgM with NaIO_4 , in conditions known to degrade serologically detectable carbohydrate¹², did not affect

its reactivity towards anti- μ antibodies as determined by RIA.

Possibility (2) was ruled out by an experiment in which equal numbers of ST4 and W231 cells were mixed before lysis and the mixed extract titrated for IgM activity after overnight incubation at room temperature. No reduction in IgM titre compared to an extract of W231 alone was found. To test the further possibility that cellular proteases could act on T-cell μ polypeptides but not on IgM, we pulse-labelled cells with ^{35}S -methionine and searched for labelled proteins after immunoprecipitation of cell extracts with anti- μ antibodies (Fig. 2). Only with the two B lymphoma lines, W231 and W279, was specifically precipitated labelled protein apparent. As little as 1% of the material present in W231 would have been detected as a specific band, but none was found in the T lymphoma cell lysates. As W231 cells contain on average 3×10^5 IgM molecules per cell and have a doubling time in culture of 12–14 h, they must synthesize at least 6,000 molecules per cell during a 15-min labelling period. ST4 cells contain 33 C_μ RNA molecules per cell compared with 95 in W231 and would be expected to synthesize 2,000 molecules in the same time period. Thus, if fewer than 60 out of 2,000 newly synthesized μ molecules remain in ST4 cells after 15 min, their half life would have to be less than 1 min, a considerably shorter time than that required for complete translation of a μ polypeptide chain.

Possibility (3), that anti- μ antibodies bind only to antigens formed by H chain-L chain association was ruled out by demonstrating that μ polypeptides, even after SDS-gel electrophoresis, could specifically bind the antibodies. Figure 3 shows the results of a filter affinity transfer experiment^{13,14} in which purified IgM (MOPC 104E) was electrophoresed in SDS and transferred to chemically activated paper to which anti- μ antibodies had been attached. Subsequent incubation with anti- μ ^{125}I -antibodies revealed the presence of a radioactive protein zone corresponding in size to μ chain in the purified IgM channel. Thus, at least a fraction of the antibody molecules can recognize μ chains not associated with L chains.

This conclusion was confirmed in RIA of extracts of an Abelson lymphoma cell line, 18–81, which synthesizes μ chains but no L chains¹¹. In these experiments (not shown), neither the shape of the titration curve nor the maximum counts per min bound with 18–81 was different from the corresponding parameters for W231, implying that L chains contributed few, if any, antigens reactive with the present antiserum.

Possibility (4) was tested by performing RIA on extracts of W231 and ST4 cells which had been subjected to SDS treatment before assay so as to dissociate protein-protein interactions. Although SDS treatment severely reduced the ability of W231 μ chain to fix anti- μ antibodies, residual antigenic activity was nonetheless clearly present in W231 but not in ST4.

Should T lymphoma cells synthesize only a fragment of normal μ chain, RIA techniques may severely underestimate or even fail to detect it, depending on the number of antigenic sites contained. In the immunoprecipitation experiments of Fig. 2, anti- μ antibodies were used in large (> 10 -fold) excess over the amount of IgM known to be present in the W231 extract, to facilitate antibody interactions with weak antigens. However, only in the two B lymphoma cell lines were μ polypeptides detected: in the T lymphoma lines, neither intact μ polypeptides nor polypeptides of altered size were detected. A single μ antigenic determinant per molecule would have resulted in specific precipitation in this test.

At face value, the data herein do not support the inference, from the existence of C_μ RNA in T lymphoma and thymus cells, that the C_μ gene may specify part of the T-cell antigen receptor^{8,9}. If C_μ RNA concentration was the sole determinant of μ polypeptide production, ST4 cells should contain about 10^5 μ molecules per cell, which would have been readily detected. Furthermore, the C_μ RNA molecules in ST4 cells contain nucleotide sequences corresponding to all four C_μ domains¹⁵ and, if translated, their protein products would be expected to express the antigenic determinants required for serological detection. Therefore, neither the qualitative characteristics (as

Fig. 2 Immunoprecipitation of detergent extracts of biosynthetically labelled lymphoma cells. 2×10^7 cells were labelled for 15 min at 37°C in 10 ml culture medium containing 10% fetal calf serum and $50 \mu\text{Ci ml}^{-1}$ ^{35}S -methionine (specific activity $\sim 1,000 \text{ Ci mmol}^{-1}$). Incubations were terminated by the addition of 1 ml $10\times$ concentrated Triton X-100 buffer and supernatants were prepared as in Fig. 1 legend. Samples of each supernatant containing equal amounts of radioactivity were reacted with $10 \mu\text{g}$ anti- μ ($\alpha\mu$) Ig, or with $10 \mu\text{g}$ normal rabbit IgG (NR) by overnight incubation at 4°C . $10 \mu\text{l}$ of packed, formalin-fixed *Staphylococcus A* organisms were added to each sample, incubated for 1 h at 4°C and then washed twice in Triton X-100 buffer containing 1% (w/v) BSA, pH 8.5 (HCl), and once in Triton X-100 buffer lacking BSA. Samples were subject to SDS electrophoresis on 12% polyacrylamide gels¹⁷, stained to visualize molecular weight marker proteins and then fluorographed¹⁸. H, L, positions of IgM heavy chain and light chain, respectively.

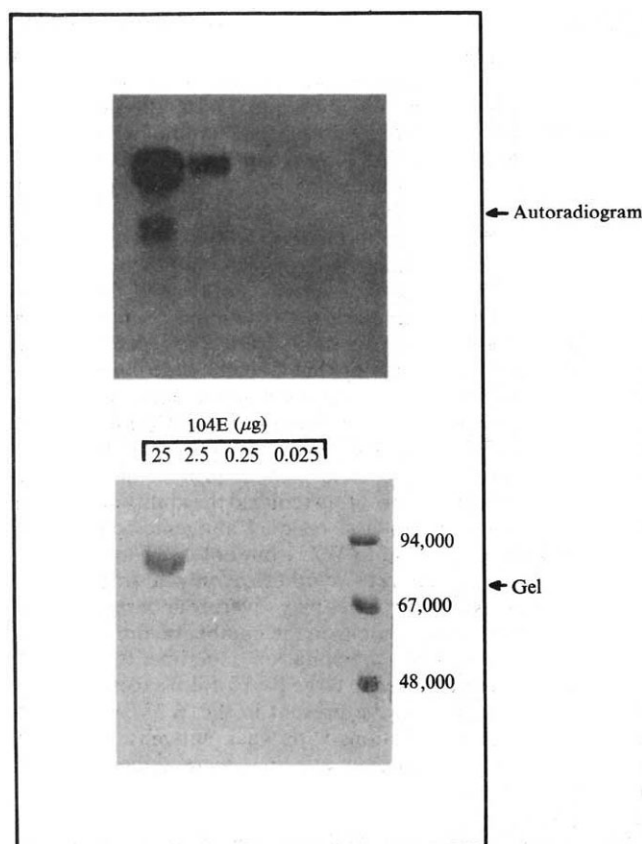
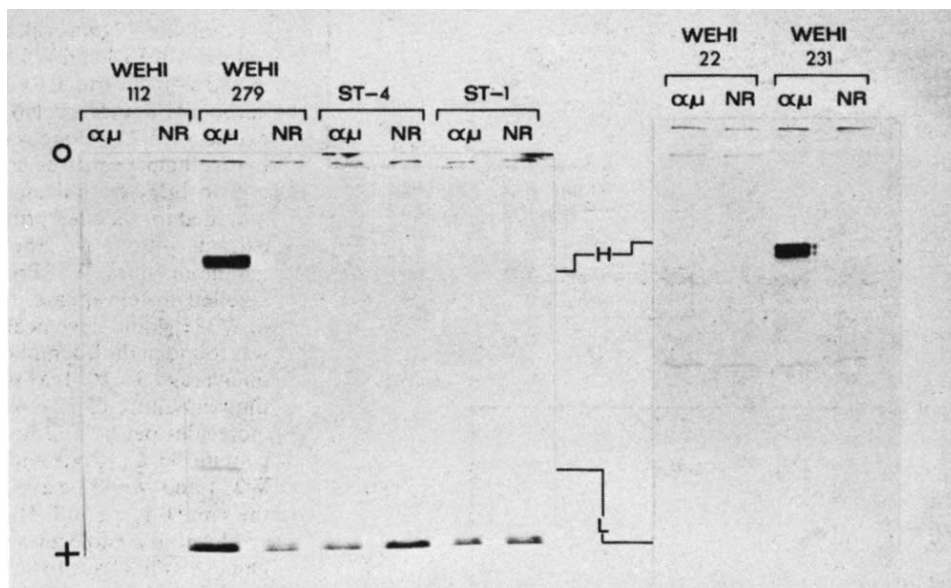


Fig. 3 μ polypeptides do not require associated L chains to react with anti- μ antibodies. The indicated amounts of purified MOPC 104E IgM and a molecular weight marker mixture of proteins containing phosphorylase *b* (94,000), BSA (68,000) and ovalbumin (48,000) were separated by SDS-gel electrophoresis. The gel was then soaked in phosphate-buffered saline (PBS) containing 1% Triton X-100 to leach out SDS. Proteins were transferred from the gel to a strip of Whatman 540 paper to which anti- μ Ig had previously been attached. After transfer, the gel was stained with Coomassie blue 250R and the paper was reacted by overnight incubation with 20 ml anti- μ - ^{125}I -Ig ($2 \times 10^6 \text{ c.p.m. ml}^{-1}$, specific activity $10 \mu\text{Ci } \mu\text{g}^{-1}$) dissolved in PBS containing 0.05% Tween 20, 1% (w/v) BSA. Filter paper was chemically activated¹⁴ and antibodies were attached as described¹³. After 'radioactive staining', the paper was washed six times with 1 l of PBS, dried and autoradiographed at -70°C using preflashed X-ray film (Kodak XRP-5) and a fast tungstate intensifying screen (Ilford). The autoradiogram (a) was obtained after a 4-h exposure and is aligned with the stained gel (b). The lower molecular-weight radioactive material beneath the μ chain band in the leftmost channel is a degradation product of 104E IgM.

acquired during later states of differentiation. If the same programme were to apply to the T-cell lineage, then μ polypeptide synthesis could be expected at a later stage of development than that represented by the T lymphoma lines of this study.

We thank Dr D. J. Kemp for making data available to us before publication and for critically reading this manuscript, and Drs K. Shortman and E. Siden for providing rabbit anti-IgM antiserum and the 18-81 cell line respectively. Sue Gherashe and Angela Apa gave excellent technical assistance.

Received 12 August; accepted 1 October 1980.

presently known) nor the relative amounts of T lymphoma μ RNA compared to its B cell (e.g. W231) counterpart provide an explanation for the absence of μ polypeptides.

It therefore seems that post-transcriptional control of μ polypeptide synthesis may operate in T lymphoma cells. Further evidence for the occurrence of untranslated C_μ RNA comes from some of our recent observations with Abelson virus-transformed pre-B cell lines. Four lines contain C_μ RNA but only one of them, 18-81, produces μ chains (I.D.W., A.W.H. and D. J. Kemp, unpublished observations). Thus, in the B lineage, the ability to translate μ RNA molecules may be

1. Eichmann, K. *Eur. J. Immun.* **4**, 296-302 (1974).
2. Cosenza, H., Julius, M. & Augustin, A. *Immun. Rev.* **34**, 3-33 (1977).
3. Eichmann, K. & Rajewsky, K. *Eur. J. Immun.* **5**, 661-666 (1975).
4. Nisonoff, A., Ju, S. T. & Owen, F. *Immun. Rev.* **34**, 89-118 (1977).
5. Bach, B. A., Sherman, L., Benacerraf, B. & Greene, M. I. *J. Immun.* **121**, 1460-1468 (1978).
6. Hirai, Y. & Nisonoff, A. *J. exp. Med.* **151**, 1213-1231 (1980).
7. Sy, M. S., Dietz, M. H., Germain, R. N., Benacerraf, B. & Greene, M. I. *J. exp. Med.* **151**, 1183-1195 (1980).
8. Kemp, D. J., Harris, A. W., Cory, S. & Adams, J. M. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2876-2880 (1980).
9. Kemp, D. J., Wilson, A., Harris, A. W. & Shortman, K. *Nature* **286**, 168-170 (1980).
10. Cory, S. & Adams, J. M. *Cell* **19**, 37-51 (1980).
11. Siden, E. J., Baltimore, D., Clark, D. & Rosenberg, N. E. *Cell* **16**, 389-396 (1979).
12. Mattes, M. J. & Steiner, L. A. *Nature* **273**, 761-763 (1978).

13. Ehrlich, H. A., Levinson, J. R., Cohen, S. N. & McDevitt, H. O. *J. biol. Chem.* **10**, 12240-12247 (1979).
14. Alwine, J., Kemp, D. J. & Stark, G. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5350-5354 (1977).
15. Kemp, D. J., Harris, A. W. & Adams, J. M. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
16. Klinman, N. R. *et al. Annls Immun. Inst. Pasteur, Paris* **127**, 489-495 (1976).
17. Laemmli, U. K. *Nature* **227**, 680-685 (1970).
18. Laskey, R. A. & Mills, A. D. *Eur. J. Biochem.* **56**, 335-341 (1975).

Methylation map of *Xenopus laevis* ribosomal RNA

B. E. H. Maden

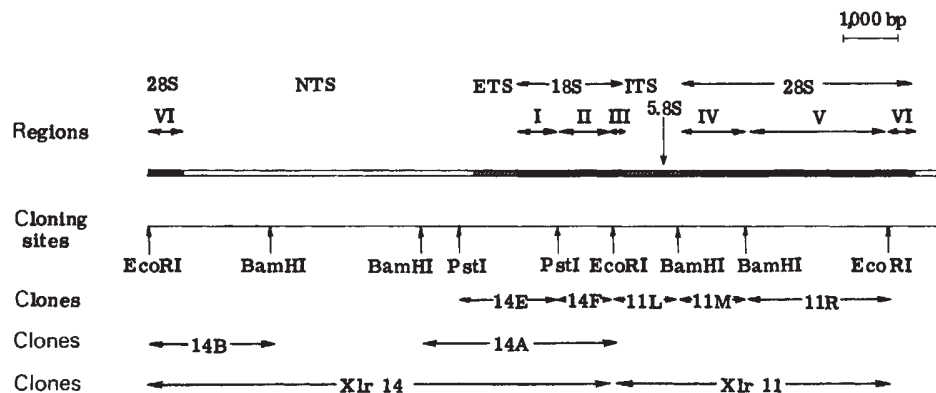
Department of Biochemistry, University of Glasgow,
Glasgow G12 8QQ, UK

One of the most enigmatic features of eukaryotic ribosomal RNA is the presence of many methylated nucleotides. The numbers of RNA methyl groups range from approximately 70 per ribosome in yeast¹ to over 100 in vertebrates^{2,3}. Here it is shown that the methylated nucleotides in *Xenopus laevis* rRNA are broadly but non-uniformly distributed. In 18S rRNA 2'-O-methylations are partly concentrated in the 5' region and base methylations near the 3' end. In 28S rRNA methyl groups are infrequent in the 5' region, moderately frequent in the central region and abundant in an 1,100-nucleotide tract near the 3' end.

To map the methylated nucleotides rRNA was radioactively labelled in *X. laevis* cultured kidney cells. 18S or 28S rRNA was then hybridized to cloned restriction fragments of rDNA (Fig. 1) or to smaller, purified restriction fragments derived from the clones (see Fig. 3, below). As only part of the rRNA was complementary to each restriction fragment, the noncomplementary RNA could be eliminated by digestion with T₁ ribonuclease. The complementary rRNA was then eluted, examined by fingerprinting, and the methylated oligonucleotides in each region identified. (Limited, partial mapping data were previously obtained by this general method^{4,5}.)

Figure 2 shows some representative fingerprints. Numerous further fingerprints were obtained covering all the regions and subregions in Fig. 3. As most methylated T₁ products occur once per molecule of 18S or 28S rRNA^{2,3}, the presence or absence of particular products in different fingerprints was usually clearly recognizable, especially when autoradiographs of fingerprints which had been run simultaneously were superimposed. Additional criteria for sequence identification were alkaline hydrolysis followed by electrophoretic analysis of methylated components^{2,3}, and use of the combined T₁ plus pancreatic ribonuclease fingerprinting system^{6,7}. This system resolves, in the form of well separated digestion products, several methylated components that occur within high-numbered T₁ products that are not well resolved in Fig. 2.

Fig. 1 Slightly more than one repeat of *X. laevis* rDNA, showing cloning sites, cloned fragments and rRNA regions. Cloned fragments X1r14 and X1r11 have been described elsewhere^{4,5,21,22}. (X1r)14B and 14A were subcloned from X1r14 into pBR322 (ref. 23). Remaining fragments were subcloned into pBR322 in this work. Cloning sites in rDNA permit division of rRNA into following principal regions: I, 18S 5' end → PstI (628); II, 18S PstI → EcoRI (~950); III, 18S EcoRI → 3' end (227); IV, 28S 5' end → BamHI (~1,100); V, 28S BamHI → EcoRI (~2,500); VI, 28S EcoRI → 3' end (500). Numbers in parentheses are numbers of RNA nucleotides, determined from sequencing, hybridization and restriction data (refs 13, 23-25 and unpublished work of M. Salim, L. Hall and B.E.H.M.). NTS, ETS, ITS denote nontranscribed, external-transcribed and internal-transcribed spacers, respectively. NTS contains additional BamHI (ref. 22) and PstI sites, not shown. bp, Base pairs.



Nearly all unimolar products mapped uniquely within a single rRNA region or subregion, and in general the relative molar recoveries were in good agreement with the yields previously determined for whole 18S or 28S rRNA. A few products hybridized to two adjacent rDNA fragments (see Table 1). In each case the most likely explanation is that the RNA methylated oligonucleotide is encoded across or very close to the respective restriction site in rDNA. There were no instances of methylated sequences cross-hybridizing to widely separated regions in rDNA. Of the few methylated oligonucleotides that occur more than once per 18S or 28S rRNA, approximately correct aggregate recoveries were obtained from the sub-region(s) in which the spots mapped.

The findings are summarized in Fig. 3 and Table 1. In 18S rRNA methyl groups occur in all subregions, but their distribution is strikingly non-uniform, especially when 2'-O-ribose methylation and base methylation are scored separately. This distinction is biologically important because during eukaryotic ribosome formation 2'-O-methyl groups are added very rapidly to ribosomal precursor RNA in the nucleolus^{2,8} probably to nascent chains⁹, whereas the few base methyl groups in 18S rRNA are added later during ribosome maturation^{2,8,10}. 2'-O-methyl groups are relatively much more abundant in the 5' part of the molecule than in the central or 3' parts. 60% of the 2'-O-methyl groups are in the 5' 40% of the molecule. In contrast, the dimethyl A sequence (product 30) is close to the 3' end^{4,11} and the m⁶A sequence (product 34) is also in region III. The m⁷G sequence (product 49) is on the boundary between regions II and III, and the sequence containing the hypermodified base, am⁵Ψ, (product 37)¹⁰ is within the rather large subregion IIb.

In 28S rRNA the distribution of methyl groups is quite different from that found in 18S rRNA. At the 5' end there is a long tract of about 900 nucleotides containing only one methylated component, Am-G. There is then a small cluster of methyl groups in region IVb and the first part of Va, followed by a 'trough' in the second part of Va. Near the middle of the molecule a group of 16 methyl groups occupies the first two parts of region Vb. There follows another, distinct trough between the first of two PvuII sites and the second PvuII site or possibly the adjacent HincII site. The next 1,100 nucleotides (regions Vc and VIa) contain by far the highest abundance and concentration of methyl groups in 28S rRNA—36 2'-O-methyl groups and four of the five base methyl groups, or 60% of the total 28S methyl groups in little over 25% of the molecule. Finally, the 200 nucleotides at the 3' end contain only one methyl group. These 28S findings confirm and much extend in detail those of Brand and Gerbi⁵.

28S rRNA possesses several sequences in which two or more methyl groups occur within a single T₁ ribonuclease oligonucleotide³. These multiply methylated sequences are distributed between the three larger-scale clusters. One such oligonucleotide occurs in region IVb (product 14, containing the

unusual, base-modified m^1A), another in the adjacent part of region Va, two more in the first part of region Vb and three in regions Vc plus VIa. The central location of the triply methylated sequence, Am-Gm-Cm-A in region Vb, is of particular interest, as an identical component was once assigned to the 5' end of mammalian 28S rRNA¹², although results from our laboratory did not support this assignment⁶.

The most striking feature of the methyl group distribution is its large-scale nonuniformity, both in 28S rRNA and to a lesser extent in 18S rRNA. Methylation requires recognition by methylases of specific sites in ribosomal precursor RNA or (for the few late methylations) in mature 18S rRNA. The recognition sites are evidently much more abundant in some regions of rRNA than in others. It seems likely that the frequencies of methylation sites in different parts of rRNA are somehow

related to other aspects of ribosome structure and function. Some indications of a possible link between rRNA methylation and biological function may be gained from the following. (1) The resolution of the methylation map is sufficient in many regions to permit accurate pinpointing of methyl groups in relation to sequence data obtained from rDNA¹³. This analysis is currently being extended for 18S rRNA with a view to definitive characterization of the methylation sites in this rRNA. (2) The rRNAs of several vertebrates yield similar fingerprints of methylated oligonucleotides to *X. laevis*, with a few distinguishing features³. It is therefore likely that the methylated sequences are similarly distributed along the primary structure of rRNAs in other vertebrates. (3) Gerbi and co-authors^{5,14} have pointed out that much of the highly methylated region near the 3' end of *Xenopus* 28S rRNA is highly conserved between

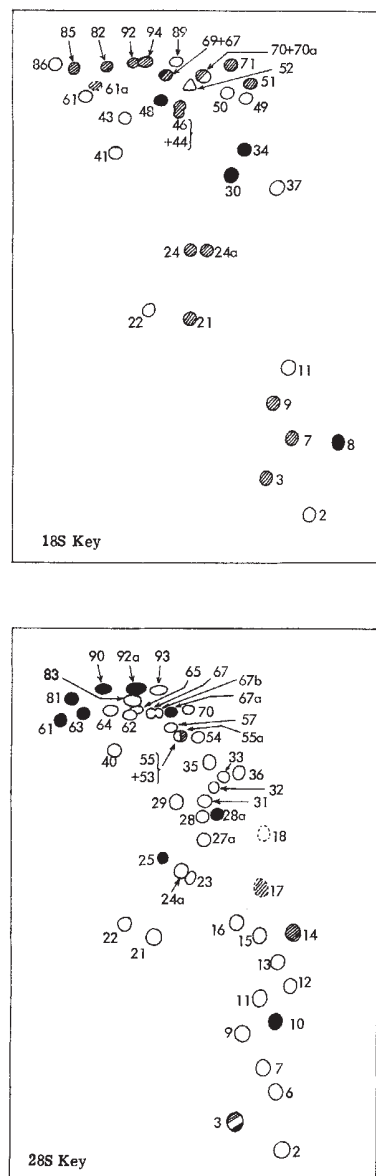


Fig. 2 T_1 ribonuclease fingerprints of regions of *X. laevis* methyl-labelled rRNA. *X. laevis* cultured kidney cells were labelled with sufficient [$Me-^{14}C$]-methionine (Amersham, ~ 50 mCi mmol⁻¹) to permit growth for 3–4 generations, and 18S and 28S rRNA were prepared as described elsewhere⁶. A slight molar excess of the appropriate rRNA was hybridized to ~ 100 μ g of linearized recombinant plasmid DNA containing fragment 14E for region I, 14F for region II and so on (see Fig. 1). Details of hybridization, trimming of hybrids with T_1 ribonuclease and recovery of T_1 -resistant, hybridized RNA are given in ref. 13. Fingerprinting was as described²⁶. First panel on upper line shows diagram of whole 18S rRNA; fingerprint has been published³. Shaded spots (3, 7, 9...) are in region I (fingerprint in second panel), open spots (2, 11...) are in region II (third panel), filled spots (8, 30...) are in region III (last panel). Spots 44 and 46 usually co-migrate³, but have separated in (I). 61a is a faint component (0.3 M) which is not included in Table 1. First panel on lower line shows diagram of whole 28S rRNA; fingerprint has been published³. Shaded spots are in region IV (second panel), open spots in region V (third panel), filled spots in region VI (last panel). Spot 3 occurs in each of the three regions. Spot 17 (region IV) is almost certainly a derivative of spot 14 resulting from isomerization of m^1A to m^6A during hybridization at elevated temperature and subsequent elution^{7,27}; individually the yields of the two spots varied but collectively they always amounted to two methyl groups (m^1A plus Am-C; Table 1). Spot 18 is recovered in low yield and is not included in Table 1. 83 in the key is a mixture of 83, 83a and 83b (Table 1); 93 in the key is a mixture of 93, 93a and probably another component, 93b, in Table 1. Products 85 upwards in 18S and 28S rRNA showed variable mobilities or streaking in T_1 fingerprints but were definitively mapped by additional criteria outlined in text.

Table 1 Methylated sequences in *X. laevis* rRNA

18S subregion	Methylated sequences	Methyl 2'-O, base	Length	Methyl 2'-O, base	Length	2'-O-methyl per 100N
Ia	46, ... Am-U-G; 87, ... Am-U ...; 85, U-Um-C-C-U-U-Um-G	4 (-)	160	13 (-)	502	2.59
Ib	3, Am-G	1 (-)	102			
Ic	7, A-Cm-G; 9, Am-A-G; 24, U-Am-G; 24a, A-Gm-G(×2); 44, ... Um-C ...; 51, ... Cm-C ...; 71, ... Am-A-A-U ...	8 (-)	240			
Id	21, Gm-G; 67, ... Um-A-A-U ...; 70, ... A-Gm-C ...; 82, ... Am-U ...; 92, ... A-A-Am-U ...	5 (-)	126	7 (-)	246	2.85
IIa	52, ... Am-A ...; 61, U-U-Gm-G	2 (-)	120			
IIb	2, Cm-G; 11, A-Am-C-G(×2); 22, Um-G; 37, ... amΨ ...; 41, Gm-U-G; 43, A-Um-U-G; 86, Um-U ...	7 (1)	~550	7 (1)	~550	1.27
IIc	50, A-Gm-G; 70a, ... Cm-C ...; 89, ... Um-C ...; 49, m ⁷ G-A-A-U ...	3 (1)	~250	6 (6)	~480	1.25
III	48, ... Am-A-G; 8, Cm-C-C-G; 69, ... Um-G; 34, ... (m ⁶ A, A)C ...; 30, m ₂ A-m ₂ A-C ...	3 (5)	227			
		Total		33 (7)	~1,800	1.83
28S subregion						
IVa	3, Am-G	1 (-)	~900	1 (-)	~900	0.1
IVb	14, (m ¹ A, A)C, AmC ...; 55a, ... Cm-U ...	2 (1)	~230	6 (1)	~480	1.25
Va Bam-Alu	21, Gm-G; 36, ... Gm-A-Am-A-G; 57, ... A-Am-C ...	4 (-)	~250			
Va Alu-Bgl	6, C-Am-G; 23, Cm-U-G(×0.5)	2 (-)	~500	2 (-)	~500	0.4
Vb Bgl-Sst(a)	3, Am-G; 22, Um-G; 62, Cm-A-Gm-U ...; 64, A-U-Cm-U ...; 93, ... Am-Gm-Cm-A ...	8 (-)	~400			
Vb Sst(a)-Pvu(a)	15, ... C-Am-G; 21, Gm-G; 27a, U-Cm-A-G; 35, A-A-Um-G; 54, ... CmC, UmC ...; 70, ... Am-U ...; 83a, ... Gm-G ...	8 (-)	~350	16 (-)	~750	2.13
Vb Pvu(a)-Hinc	21, Gm-G(×0 or 1)	0-1 (-)	~250	0-1(-)	~250	<0.5
Vc Hinc-Sst(b)	2, Cm-G; 7, A-Cm-G; 9, A-Am-G; 11, A-Am-C-G; 12, ... Cm-A-G; 13, Cm-C ...; 16, C-C-Gm-G; 21, Gm-G(×2 or 3); 28, ... (m ⁶ A, A)C ...; 29, A-A-Ψm-G; 31, A(A, Gm-A)G; 32, ... A-Am-A-G; 33, mC, A-A-Am-U; 53, Gm-U ...; 55, ... Um-C ...; 65, ... A-Am-C ...; 67b, ... Am-U ...; 83, ... Um-Gm-U ...; 93a, ... A-Cm-U ...; 93b, ... Cm-C ... (?)	21-22 (2)	~580	~35 (4)	~1,100	3.18
Vc Sst(b)-EcoRI	21, Gm-G; 24a, A-Gm-G; 40, Um-A-U-G; 67, ... mC ...; 83b, ... A-Gm-C ...	4 (1)	~200			
Vla	3, Am-G; 10, A-Cm-C-G; 25, Am-U-G(×0.5); 61, U-U-Gm-G; 63, m ³ U-U ...; 67a, ... A-Am-U ... (0.5); 81, ... Gm-U-G; 90, ... Um-C, (A, Gm-A)C; 92a, ... Um-Gm-Ψ ...	10 (1)	300			
Vlb	28a, ... Cm-U ...	1 (-)	200	1 (-)	200	0.5
		Total		~62 (5)	~4,150	1.49

The methylated sequences shown here are as reported in ref. 3, although the two low-yield products, 61a (18S) and 18 (28S), are not shown as they have not been precisely mapped. Most sequences longer than tetranucleotides are given only in part. Further details, where available, are in refs 3 and 13. The nomenclature (revised) is as follows: The former 18S spot 70 (ref. 3) has two components, now called 70 (region Id) and 70a (region IIc). Spot 89 is correct identification for a spot previously³ designated 87. Former 28S spot 67 has two components now called 67 and 67b (both in region Vc). Identification and mapping of product 93b are tentative. In regions Ia, IIa and III, products are listed in the order in which they appear in the sequence. In other regions they are listed in arbitrary order. In the last columns data are grouped to highlight main features of methyl group distribution, firstly by subregions, then by larger rRNA segments with differing methylation densities. Further details on experimental methods are given in Fig. 3 legend.

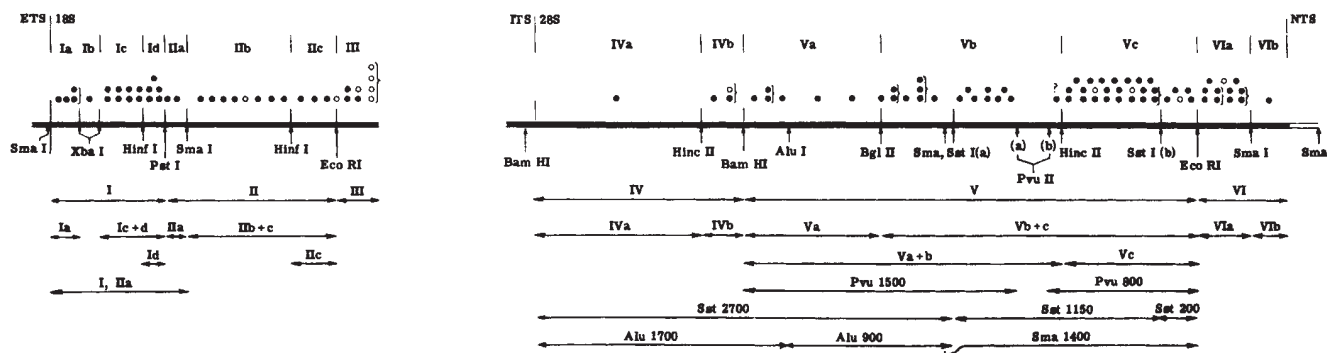


Fig. 3 The general features of methyl group distribution and restriction fragments used. Restriction map is to scale and methyl groups were reliably assigned to each subregion, but precise locations within each subregion are so far known in only a few instances^{11,13}. Closed circles denote 2'-O-ribose methyl groups, open circles denote base methyl groups, brackets denote multiply methylated oligonucleotides. Preparative hybridization required rDNA fragments at least a few hundred nucleotides long; shorter fragments did not adhere to nitrocellulose filters. There was no upper limit to useful fragment length, and presence of vector DNA (which adjoined many of the fragments in Fig. 3) did not cause spurious hybridization. As large quantities of restriction fragments were required (equivalent to ~100 µg of whole plasmid per [Me-¹⁴C] fingerprint) separations were usually on sucrose gradients (10–25%, in 1 M NaCl, 0.001 M EDTA, 0.025 M Tris-HCl, pH 8); restriction enzymes which cut infrequently were chosen where possible, and several subregions were mapped by 'overlaps' (lower part of figure). Restriction digests were planned from published maps of rDNA^{24,25}, unpublished data (below), map of pBR322 (ref. 28) and known orientation of rDNA insert. In the 18S gene, sites shown for *Pst*I, *Sma*I and *Eco*RI are only sites for these enzymes. There are several *Hinf*I sites; those shown are next left of *Pst*I and *Eco*RI sites respectively²⁵. *Hinf*I digestion of appropriate plasmids yielded these short rDNA regions attached to an easily separated fragment of pBR322 DNA. In the 28S gene, sites shown for *Hinc*II, *Bam*HI, *Bgl*II, *Sst*I, *Pvu*II and *Eco*RI are only sites for these enzymes. There are several *Sma*I sites^{24,25}; those shown are the last two in the gene. There are also several *Alu*I sites^{24,25}; the one shown is first in 28S gene. The *Sst*I and *Pvu*II sites are subsets of *Alu* sites and were mapped in this work. They are the second and sixth or seventh *Alu* sites (*Sst*I) and the third and fourth *Alu* sites (*Pvu*II). Left-hand sites defining *Sst*2,700 and *Alu*1,700 are in ITS and 5.8S gene respectively in *Xlr*11. See ref. 24 for *Alu* map. After purification of rRNA fragments by preparation hybridization, most products were recovered in ~ unimolar yield in T₁ ribonuclease fingerprints. Those marked (×2) or (×0.5) in Table 1 occur in approximately these frequencies. (0.5 M spots were treated as unimolar in the quantitative summary). Spot 37 and some of highest-numbered spots showed variable mobilities and yields in T₁ fingerprints. Smaller 'derivative' products occurred in good yields in combined T₁ plus pancreatic ribonuclease fingerprints. See ref. 7 for correlation between the two fingerprinting systems. A few products hybridized weakly to two adjacent fragments. In each case the probable explanation is that the methylated sequence occurs close to a restriction site: spot 52 hybridized with region IIa and weakly with region Id. Sequence data (M. Salim and B.E.H.M., unpublished) place this sequence as next product to right of *Pst*I site. Spot 49 hybridized with regions IIc and III. Data on RNA oligonucleotide are consistent with DNA sequence encompassing *Eco*RI site (M. Salim, L. Hall and B.E.H.M., unpublished). Spot 64 in 28S RNA hybridizes with region Vb and weakly with Va. Sequence G-A-U-Cm-U-U-G probably overlaps *Bgl*II site, (A)-G-A-T-C-T ... One of the Gm-G residues in 28S RNA is probably close to second *Hinc*II site (demarcating regions Vb and Vc); it is not yet known on which side.

Xenopus and yeast, whereas the almost unmethylated region near the 5' end shows little sequence conservation between these distantly related eukaryotes. It will be extremely useful to determine at the level of nucleotide-sequence analysis the precise relationship between sequence conservation and secondary modification sites during eukaryotic rRNA evolution. (4) The only region of clear sequence homology so far identified between *Xenopus* 18S rRNA and *Escherichia coli* 16S rRNA is a short region at the 3' end^{11,14}. This short, homologous region contains the dimethyl A sequence. The 5' one-third of *E. coli* 16S rRNA is completely unmethylated^{15,16} in contrast to the highly 2'-O-methylated 5' one-third of *Xenopus* 18S rRNA.

Received 14 July; accepted 18 September 1980.

1. Klootwijk, J. & Planta, R. J. *Eur. J. Biochem.* **39**, 325-333 (1973).
2. Maden, B. E. H. & Salim, M. J. *molec. Biol.* **88**, 133-164 (1974).
3. Khan, M. S. N., Salim, M. & Maden, B. E. H. *Biochem. J.* **169**, 531-542 (1978).
4. Maden, B. E. H. & Reeder, R. H. *Nucleic Acids Res.* **6**, 817-830 (1979).
5. Brand, R. C. & Gerbi, S. A. *Nucleic Acids Res.* **7**, 1497-1511 (1979).
6. Khan, M. S. N. & Maden, B. E. H. *J. molec. Biol.* **101**, 235-254 (1976).
7. Maden, B. E. H. & Khan, M. S. N. *Biochem. J.* **167**, 211-221 (1977).
8. Brand, R. C., Klootwijk, J., Van Steenberg, T. J. M., De Kok, A. J. & Planta, R. J. *Eur. J. Biochem.* **75**, 311-318 (1977).
9. Greenberg, M. & Penman, S. *J. molec. Biol.* **21**, 527-535 (1966).
10. Brand, R. C., Klootwijk, J., Planta, R. J. & Maden, B. E. H. *Biochem. J.* **169**, 71-77 (1978).
11. Hagenbüchle, O., Santer, M., Steitz, J. A. & Mans, R. J. *Cell* **13**, 551-563 (1978).
12. Choi, Y. C. & Busch, H. *Biochem. biophys. Res. Commun.* **58**, 674-682 (1974).
13. Salim, M. & Maden, B. E. H. *Nucleic Acids Res.* **8**, 2871-2884 (1980).
14. Gourse, R. L. & Gerbi, S. A. *J. molec. Biol.* **140**, 321-339 (1980).
15. Carbon, P., Ehresmann, C., Ehresmann, B. & Ebel, J. P. *FEBS Lett.* **94**, 152-156 (1978).
16. Brosius, J., Palmer, M. L., Poindexter, J. K. & Noller, H. F. *Proc. natn. Acad. Sci. U.S.A.* **75**, 4801-4805 (1978).
17. Muto, A., Ehresmann, C., Fellner, P. & Zimmermann, R. A. *J. molec. Biol.* **86**, 411-432 (1976).
18. Cole, M. D., Beer, M., Koller, Th., Stycharz, W. A. & Nomura, M. *Proc. natn. Acad. Sci. U.S.A.* **75**, 270-274 (1978).
19. Newberry, V., Yaguchi, M. & Garrett, R. A. *Eur. J. Biochem.* **76**, 51-61 (1977).
20. Ehresmann, C., Stiegler, P., Carbon, P., Ungewickell, E. & Garrett, R. A. *Eur. J. Biochem.* **103**, 439-446 (1980).
21. Dawid, I. B. & Wellauer, P. K. *Cell* **8**, 443-448 (1976).
22. Botchan, P., Reeder, R. H. & Dawid, I. B. *Cell* **11**, 599-607 (1977).
23. Sollner-Webb, B. & Reeder, R. H. *Cell* **18**, 485-499 (1979).
24. Boseley, P. G., Tuyns, A. & Birnstiel, M. L. *Nucleic Acids Res.* **5**, 1121-1137 (1978).
25. Boseley, P. G., Moss, M., Mächler, M., Portmann, R. & Birnstiel, M. L. *Cell* **17**, 19-31 (1979).
26. Brownlee, G. G. *Determination of Sequences in RNA* (North Holland, Amsterdam, 1972).
27. Macon, J. B. & Wolfenden, R. *Biochemistry* **7**, 3453-3458 (1968).
28. Sutcliffe, J. G. *Nucleic Acids Res.* **5**, 2721-2728 (1978).

Structure of crystalline actin sheets

Ueli Aepli, Phillip Ross Smith*, Gerhard Isenberg & Thomas D. Pollard

Department of Cell Biology and Anatomy, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205

* Department of Cell Biology, New York University Medical Center, New York, New York 10016

Although actin is one of the most abundant proteins found in nature, little detailed information about its molecular structure is available beyond the amino acid sequence^{1,2}. Electron microscopy of negatively stained filaments combined with three-dimensional image reconstruction techniques have revealed the overall size and shape of the actin monomer at 25 Å resolution³⁻⁵. Higher resolution structural data can be expected from electron microscopy of two-dimensional crystalline arrays^{6,7} and X-ray diffraction analysis of three-dimensional crystals⁸⁻¹¹, but only very preliminary results have been reported so far. The original finding by Dos Remedios and Dickens⁶ was that skeletal muscle actin forms microcrystals and tubes in the presence of the trivalent lanthanide gadolinium (Gd³⁺). We have modified and refined their conditions to obtain large crystalline sheets of *Acanthamoeba* actin and present here a model of the actin monomer in projection to 15 Å resolution. We have found that, depending on the ionic strength used, these sheets occur in three different forms: 'cylinders', 'square type' sheets and 'rectangular type' sheets. These different polymorphic forms are built from the same fundamental two-dimensional crystalline actin lattice, which we call the 'basic sheet'. The present concerns the structural analysis of these basic sheets; the crystal polymorphism will be discussed in detail elsewhere (U.A. *et al.*, in preparation). Furthermore, in addition to demonstrating that actin is an elongated globular molecule with a pronounced asymmetric shape in and perpendicular to the plane of the sheet, our results indicate that these crystalline actin sheets might be suitable for three-dimensional structure determination by low-dose electron microscopy of unstained specimens^{12,13} to at least 10 Å resolution.

An electron micrograph of a negatively stained basic sheet from an opened-up cylinder (Fig. 1A) reveals that these sheets

This region of *E. coli* 16S rRNA plays a crucial role in ribosome assembly through interaction with ribosomal protein S4 (refs 17-20). As yet the function of the 5' one-third of eukaryotic 18S rRNA is unknown.

I started this work at the Carnegie Institution of Washington, Department of Embryology, Baltimore. I thank D. D. Brown, S. McKnight, R. Peterson, R. Reeder and Barbara Sollner-Webb for discussion, advice on cloning and gifts of restriction enzymes. I also thank J. Forbes and Mary Robertson for technical assistance in Glasgow, and J. Sheddon for discovery of the *Sst*I sites. The work was supported by the MRC.

are made of a near-rectangular lattice with average lattice parameters $a = 56.5$ Å, $b = 65.5$ Å and $\gamma = 85.5^\circ$ (average of 60 sheets). Figure 1B displays an optical diffraction pattern recorded from a circular area of the micrograph including about 1,000 56.5×65.5 Å unit cells. The fact that strong $(33 \text{ Å})^{-1}$ second order reflections are visible along b' , with the $(65.5 \text{ Å})^{-1}$ first order reflections being almost completely absent, is consistent with the observation of a pronounced 33 Å pseudo-repeat along b in the images (see Fig. 1A), indicating that there might be two molecules per unit cell, related to one another by a two-fold axis of symmetry. Optical diffraction analysis of basic sheets indicates that the long-range order can extend to 5,000 Å in both directions within the plane of the sheet, with sharp diffraction spots extending out to $(15 \text{ Å})^{-1}$ (see Fig. 1B for example $(\pm 1, \pm 4)$ reflections). We have found right- and left-handed basic sheet lattices with about equal frequencies with negatively stained specimens. However, the sheets appearing left-handed on the grid (see Fig. 1A) invariably showed better structural preservation and higher resolution detail. As is illustrated in Fig. 2, the handedness of the basic sheet observed on the grid depends on which of the two sheet surfaces is in contact with the carbon support film on adsorption.

Basic sheets prepared for electron microscopy by adsorption freeze-drying and shadowing^{14,15} showed two surface topographies. One surface was smooth and regular, with a near-rectangular right-handed lattice texture (Fig. 2A) and average lattice parameters $a = 55$ Å, $b = 63.5$ Å and $\gamma = 86.5^\circ$ (average of 40 regular-looking sheets). The other surface appeared coarse and irregular, with no indication of a lattice texture, and showed a granularity similar to that found in the background (Fig. 2B). Occasionally, basic sheets folded over, allowing the two distinct surface topographies to be seen simultaneously (Fig. 2C). Optical diffraction patterns recorded from the regular surface of shadowed basic sheets yielded the $(55 \text{ Å})^{-1}$ and $(63.5 \text{ Å})^{-1}$ first order reflections along a' and b' , respectively, with no indication of the $(32 \text{ Å})^{-1}$ second order reflections along b' observed with negatively stained sheets (Fig. 1B). This must be due to the inability to resolve the two molecules in the unit cell by our freeze-drying and shadowing technique. In contrast, basic sheets with the irregular surface topography did not produce any reciprocal lattice reflections on optical diffraction. From the lengths of the shadows cast at the edges of the sheets we calculated their average thickness to be 45 Å (average of

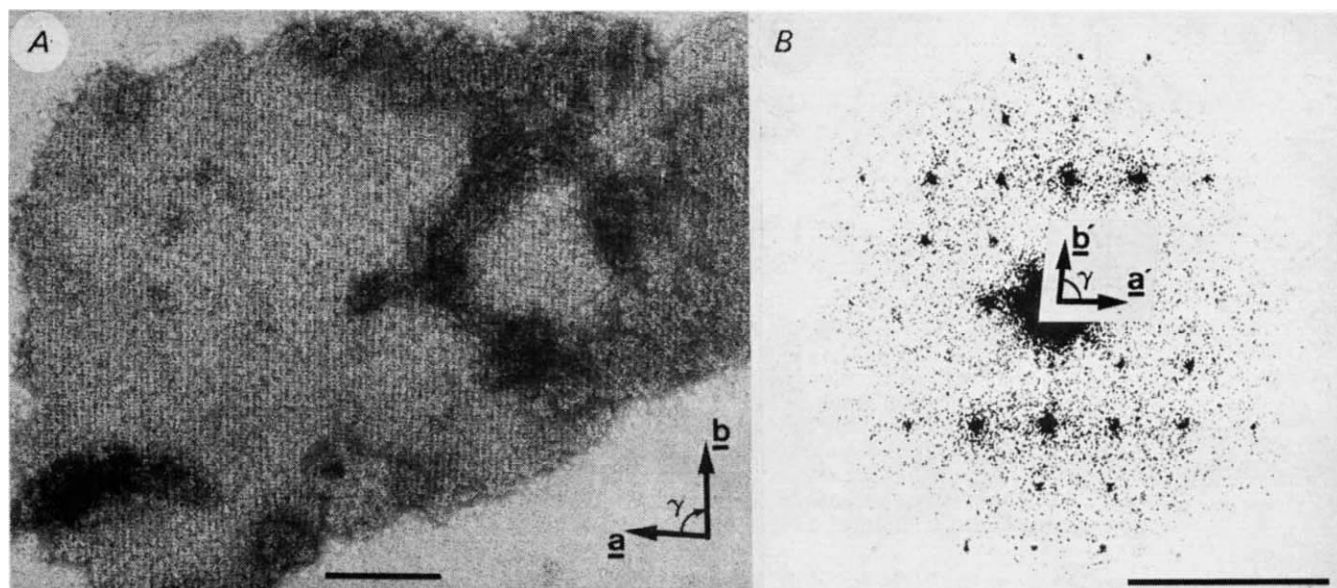


Fig. 1 A, Electron micrograph of a negatively stained actin basic sheet arising from an opened-up cylinder. Basic sheets were prepared as follows: *Acanthamoeba* G-actin, purified by a modification of the methods described^{16,17}, was diluted to 1.5 mg ml^{-1} and subsequently dialysed against 2.5 mM imidazole, pH 7.25 (4 °C), 0.25 mM dithiothreitol (DTT) and 0.005% NaN_3 for 5 h at 4 °C. The sample was then spun in a Beckman Airfuge for 15 min at 130,000g to remove aggregated material and the supernatant dialysed against 2.5 mM PIPES, pH 7.00 (4 °C), 150 mM KCl, 0.25 mM DTT, 0.005% NaN_3 and GdCl_3 (molar ratio Gd^{3+} : actin $\geq 6:1$) for 15 h at 4 °C. At this stage the sample consisted of a mixture of filaments (30–50% of aggregated material), rectangular and square type sheets, and cylinders (U.A. *et al.*, in preparation). A low-speed pellet (5,000g for 5 min) of this material was resuspended in one-third volume assembly buffer (see above) and immediately dialysed against 2.5 mM PIPES, pH 7.00, and 0.005% NaN_3 for at least 12 h at 4 °C. The sample was then spun for 10 min at 5,000g and the pellet resuspended in the same volume 2.5 mM PIPES, pH 7.00 (4 °C), and 0.005% NaN_3 . On dilution of this material to $\leq 0.75 \text{ mg ml}^{-1}$ with the same buffer, basic sheets started peeling off the rectangular and square type sheets, and the cylinders opened up into basic sheets when diluted material was left for a few hours at 4 °C. For electron microscopy, 2- μl drops were adsorbed for 60 s to carbon-coated 400-mesh per inch copper grids which were rendered hydrophilic by glow discharge in a reduced atmosphere of air. The grids were then washed for 30 s on several drops of distilled water and stained for 30 s on several drops of 0.75% uranyl formate, pH 4.25. Electron micrographs of these specimens were recorded on Kodak 4463 Electron Image Film in a Zeiss EM 10A electron microscope at $\times 50,000$ nominal magnification and developed for 4 min in 1:2 times diluted Kodak D-19 developer. To minimize the accumulated electron dose on the areas to be photographed, a beam rocking technique¹⁸ was used. The example shows a basic sheet with its regular surface in contact with the carbon support (see text and Fig. 3), and therefore its near-rectangular lattice appears left-handed (indicated by the lattice vectors **a** and **b**). Scale bar, 1,000 Å. B, Optical diffraction pattern recorded from a circular area of the electron micrograph shown in A including about 1,000 unit cells. The reciprocal lattice vectors corresponding to the $56.5 \times 65.5 \text{ Å}$ near-rectangular lattice have been drawn to scale in the central inset. Scale bar, $(20 \text{ Å})^{-1}$.

24 sheets), suggesting that the basic sheets are spanned by one actin molecule. The two distinct surface topographies of the basic sheets reflect two things: (1) the actin monomer must have a very asymmetric shape in the direction perpendicular to the plane of the sheet, and (2) all actin molecules are packed with the same polarity perpendicular to the plane of the sheet.

We have computer-filtered several micrograph areas containing well preserved negatively stained basic sheets, and have obtained the quaternary structure of the actin molecule within the sheet, and its overall size and shape in projection to 15 Å resolution. The average unit cell morphology obtained from five different $12 \times 10 \text{ } 56.5 \times 65.5 \text{ Å}$ unit cell areas chosen

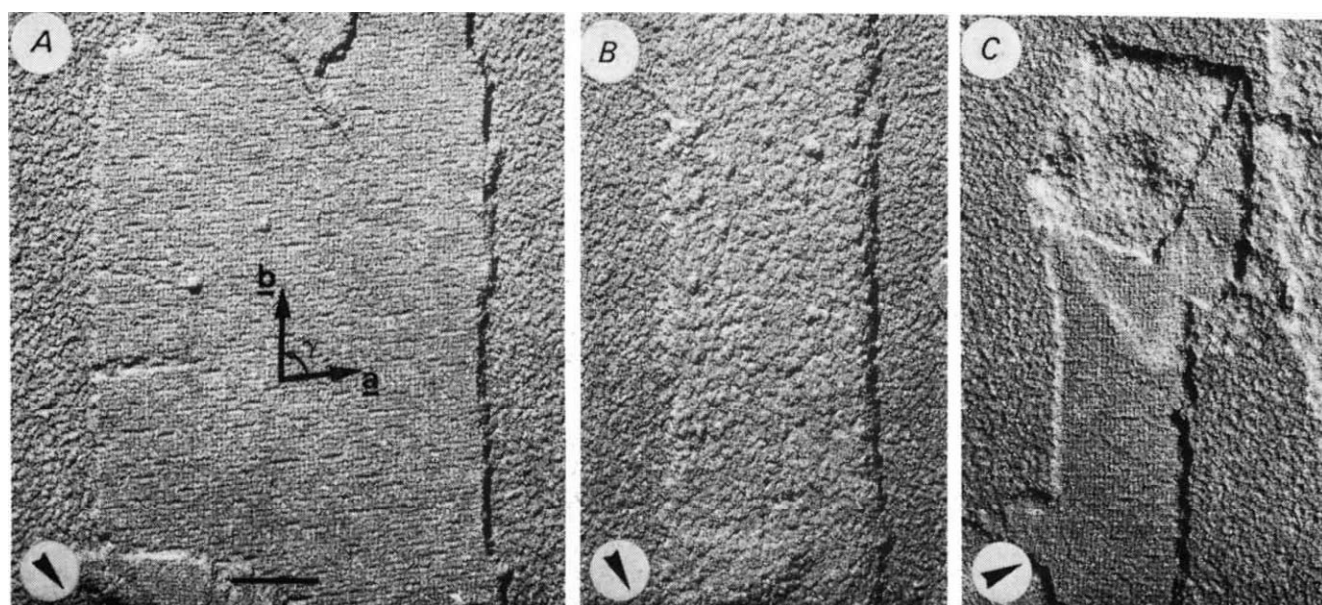


Fig. 2 Electron micrographs of freeze-dried and shadowed basic sheets exhibiting the two different surface topographies. Samples of unfixed material were prepared for electron microscopy by the adsorption freeze-drying technique described elsewhere^{14,15}. Unidirectional shadowing of the specimens was done with platinum from an electron gun at an elevation angle of about 30°. A, Example of a sheet displaying a regular, right-handed near-rectangular lattice surface (as indicated by the lattice vectors). B, Example of a sheet showing an irregular surface with no indication of a periodic lattice. C, Example of a folded-over sheet simultaneously exposing its two surfaces. The arrows in A, B and C indicate the direction of metal shadowing. Scale bar, 1,000 Å (for A, B and C).

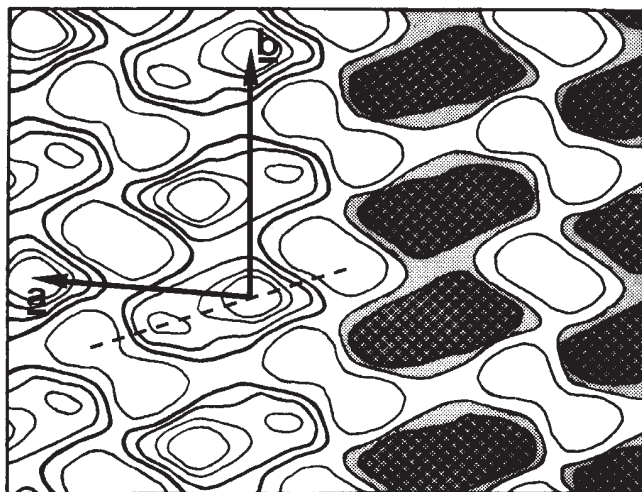


Fig. 3 A contour plot showing the average two-fold symmetrized unit cell morphology obtained from five different electron micrograph areas of computer-filtered negatively stained actin basic sheets. All five sheets selected for filtering were left-handed (see text and Figs 1, 2). The electron micrographs were digitized with an Optronics P-1000 microdensitometer using a 50- μ m square raster, from photographic enlargements onto Kodak 4147 Plus-X pan film to give an effective sampling rate of 3.25 Å. Computer filtering was performed essentially as described previously^{19,20} using $12 \times 10 \times 56.5 \times 65.5$ Å near-rectangular unit cell sheet areas. The filtered unit cells were two-fold symmetrized after searching for the best symmetry centre^{19,21}; the power loss on two-fold symmetrization averaged to 2.1% over the five samples. The filtered and two-fold symmetrized unit cells from the five samples were then aligned with respect to their two-fold centres of symmetry and summed to give the average unit cell morphology shown. The near-rectangular unit cell (as indicated by the lattice vectors **a** and **b**) contains two actin molecules related to one another by P2 symmetry. Actin dimers and monomers are indicated by the light and dark shaded regions, respectively, in the right-hand side of the figure. The broken line marks the longer axis of the actin monomer, and is defined through the major and minor peak of the subunit in projection. Scale bar, 50 Å.

from four different micrographs is shown in Fig. 3. The following reproducible features were observed. (1) The 56.5×65.5 Å near rectangular unit cell contains two actin molecules which are related to each other by a two-fold axis of symmetry perpendicular to the plane of the sheet (space group P2). (2) The actin monomer is elongated and measures about 56×33 Å in projection, with its longer axis (indicated by a broken line) being tilted by about 23° with respect to the lattice vector **a**. (3) The projected mass density of the actin subunit possesses a pronounced asymmetry with a major and a minor peak along its longer axis. (4) The negative stain accumulates as a distinctive 2-and-1 pattern between columns of actin dimers.

The unit cell parameters of the basic sheets are in agreement with those determined previously by optical diffraction of negatively stained skeletal muscle actin tubes⁷. The overall dimensions of the actin monomer are similar to those obtained from three-dimensional reconstructions of average actin filaments computed by deconvolution from filament paracrystals³. Moreover, the distinct asymmetric shape of the subunit within and perpendicular to the plane of the basic sheet is consistent with the polarity of the actin monomer found by three-dimensional reconstruction of individual actin filaments⁵. A more quantitative comparison with the actin filament structure must await completion of a three-dimensional map of the molecule computed from tilted views of the crystalline sheets. However, the quaternary structure of the actin molecules in the basic sheet differs substantially from their packing in the filament. Although it is possible that the head-to-tail bonds between actin molecules are similar in the sheets (for example along **a**) and the filaments (for example along a single strand), the lateral associations cannot be the same in the two cases. In the sheets, adjacent rows

of subunits have opposite polarity, whereas in filaments the two strands have the same polarity with respect to the filament axis. Consequently, the assembly of the basic sheets cannot be explained as an 'unwinding' of the double-helical filaments into linear double strands, followed by a side-by-side association of these linear double strands.

From the size and long-range order of the crystalline actin sheets described here, we are confident that a three-dimensional model of the actin molecule can be determined to at least 10 Å resolution by low-dose electron microscopy and three-dimensional reconstruction of tilted views of unstained specimens^{12,13}. Finally, it might be possible to decorate basic sheets stoichiometrically with actin-interacting proteins such as myosin S-1, troponin, gelation factors, profilin, DNase I and 5'-nucleotidase, and thus to map the binding sites for these molecules on the three-dimensional model of actin.

We thank Dr V. G. Allfrey for allowing us to use his Optronics P-1000 microdensitometer. This work was supported by National Institute of General Medical Sciences grants GM 27765 to U.A., GM26723 to P.R.S. and GM26338 to T.D.P. G.I. is supported by a Heisenberg Fellowship from the Deutsche Forschungsgemeinschaft.

Received 11 August; accepted 30 September 1980.

- Collins, J. H. & Elzinga, M. *J. biol. Chem.* **250**, 5915–5920 (1975).
- Vandekerckhove, J. & Weber, K. *Nature* **276**, 720–721 (1978).
- Moore, P. B., Huxley, H. E. & DeRosier, D. J. *J. molec. Biol.* **50**, 279–295 (1970).
- Spudich, J. A., Huxley, H. E. & Finch, J. T. *J. molec. Biol.* **72**, 619–632 (1972).
- Wakabayashi, T., Huxley, H. E., Amos, L. A. & Klug, A. *J. molec. Biol.* **93**, 477–497 (1975).
- Dos Remedios, C. G. & Dickens, M. J. *Nature* **276**, 731–733 (1978).
- Dickens, M. J. *Proc. R. microsc. Soc.* **13**, 80–81 (1978).
- Carlsson, L. *et al. J. molec. Biol.* **105**, 353–366 (1976).
- Mannherz, H. G., Kabsch, W. & Leberman, R. *FEBS Lett.* **73**, 141–143 (1977).
- Oriol, C., Dubord, C. & Landon, F. *FEBS Lett.* **73**, 89–91 (1977).
- Sugino, H. *et al. J. Biochem.* **86**, 257–260 (1979).
- Unwin, P. N. T. & Henderson, R. *J. molec. Biol.* **74**, 425–440 (1975).
- Henderson, R. & Unwin, P. N. T. *Nature* **257**, 28–32 (1975).
- Kistler, J., Aebi, U. & Kellenberger, E. *J. ultrastruct. Res.* **59**, 76–86 (1977).
- Smith, P. R. *J. ultrastruct. Res.* **72**, 380–384 (1980).
- Pollard, T. D., Stafford, W. F. & Porter, M. E. *J. biol. Chem.* **253**, 4798–4808 (1978).
- Gordon, D. J., Eisenberg, E. & Korn, E. D. *J. biol. Chem.* **251**, 4778–4786 (1976).
- Williams, R. C. & Fisher, H. W. *J. molec. Biol.* **52**, 121–123 (1970).
- Aebi, U., Smith, P. R., Dubochet, J., Henry, C. & Kellenberger, E. *J. supramolec. Struct.* **1**, 498–522 (1973).
- Smith, P. R. *Ultramicroscopy* **3**, 153–160 (1978).
- Aebi, U. *et al. J. molec. Biol.* **130**, 255–272 (1979).

Helix packing and subunit conformation in horse spleen apoferritin

George A. Clegg, Robert F. D. Stansfield, Philip E. Bourne & Pauline M. Harrison

Department of Biochemistry, The University, Sheffield S10 2TN, UK

An electron density map of horse spleen apoferritin at 0.28-nm (2.8 Å) resolution and its preliminary interpretation have been described previously¹. Rigorous examination of this and newer maps at the same nominal resolution but calculated from more extensive data sets, including model building in a Richards' comparator², now allows us to report on structural features in more detail. We list inter-helical angles within and between neighbouring subunits, and describe a new short region of inter-subunit anti-parallel pleated sheet. A short section of electron density not properly accounted for in the first model has also been found. We also report on two alternative ways of connecting helical regions which account almost equally well for the observed electron density, and we assess these two alternative conformations and compare them with the conformations of other known proteins.

The iron storage molecule ferritin consists of a protein shell comprising 24 subunits in 432 symmetry surrounding an inorganic 'core' of hydrated ferric oxide-phosphate. Horse

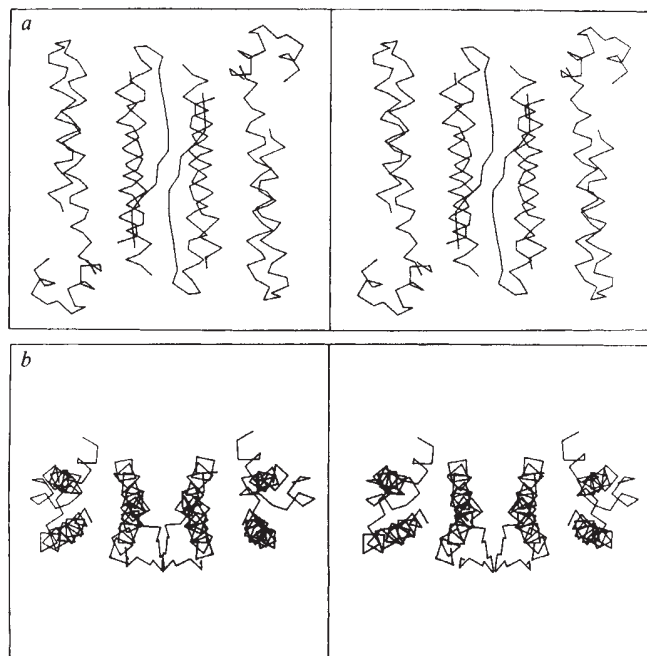


Fig. 1 Stereoscopic views of connected α -carbon coordinates for well established regions in an apoferritin dimer. *a*, The view down a molecular diad from outside the molecule. The four major helices A, B, C, D in the leftmost subunit occur: above right, below right, above left and below left, respectively. *b*, 'End-on' view perpendicular to the diad and parallel to a molecular triad. The four major helices A, B, C, D in the leftmost subunit occur: below right, above right, below left and above left, respectively. The atomic positions used here and in Fig. 2 have regular geometry and are faithful to the electron density map, but may not be correct in every detail.

spleen apoferritin (subunit molecular weight 19,000) crystallizes from CdSO_4 solution in space group F432 with $a = 18.40$ nm (184.0 Å). The asymmetric unit contains a single subunit. Electron density maps show clearly that each subunit consists of a bundle of four long helices lying parallel or anti-parallel to one another, together with two much shorter helices and stretches of irregular conformation. It is in the inter-helical connecting regions that the electron density is weakest and where difficulties are experienced in tracing the course of the polypeptide chain unambiguously. Preliminary sets of α -carbon and carbonyl oxygen atomic coordinates have been measured in the Richards' comparator and used to derive regularized sets of coordinates for main chain and β -carbon atoms.

We will first describe the regions which we consider to be reasonably well established. These are the long helices A, B, C and D, containing approximately 27, 25, 28 and 20 amino acid residues respectively, and two short helices, P, 7 residues (formerly described as a helical turn), and E, comprising about 10 residues. The total number of helical residues (117) accounts for about 68% of the main chain density. Helix B uniquely seems to contain an irregularity near the middle of its length, and N-terminal regions of helices C and D seem to be fitted best by one to two turns of 3.0_{10} helix. There is also good density for the sharp turn connecting C and D, for the long loop L and (less so) for the connecting regions between D-P to E. The directions of the helices are confirmed to be those previously assigned, and angles between pairs of helices are given in Table 1. Among the four major helices of one subunit, adjacent anti-parallel pairs lie within the range $166 \pm 5^\circ$ to one another, and diagonally opposite parallel pairs are inclined at $\sim 17^\circ$. The four helices are thus related by 222 pseudo-symmetry with pseudo-diads passing between helical pairs A, C and B, D and between C, D and A, B and a third lying nearly parallel to the long axes of the helices (and this is the long axis of the subunit). In the apoferritin 432 quaternary structure, pairs of subunits related by molecular

diads lie with their long axes anti-parallel. Angles relating all pairs of helices within this 'dimer' are given in Table 1. Interestingly, angles relating neighbouring anti-parallel helices A, A' and B, B' related by the molecular two-fold axis are similar to those within the subunit, and the same is true of the diagonally related pairs A, B' and B, A'. Thus, a complete system of eight helices is formed comprising two layers, each of which contains four anti-parallel helices, and this system itself has pseudo-222 symmetry.

In Fig. 1 stereo views of the helices and other well established regions within the dimer are shown as connected α -carbon diagrams down the diad (Fig. 1*a*) and approximately end-on to the major helices, perpendicular to the diad and parallel to a molecular three-fold axis (Fig. 1*b*). It can be seen in Fig. 1*b* that the four major helices within a single subunit together show a left-handed twist when viewed down their lengths, whereas the two pairs of A and B helices near the diad are arranged so that diagonally related pairs A, B' and B, A' converge at opposite

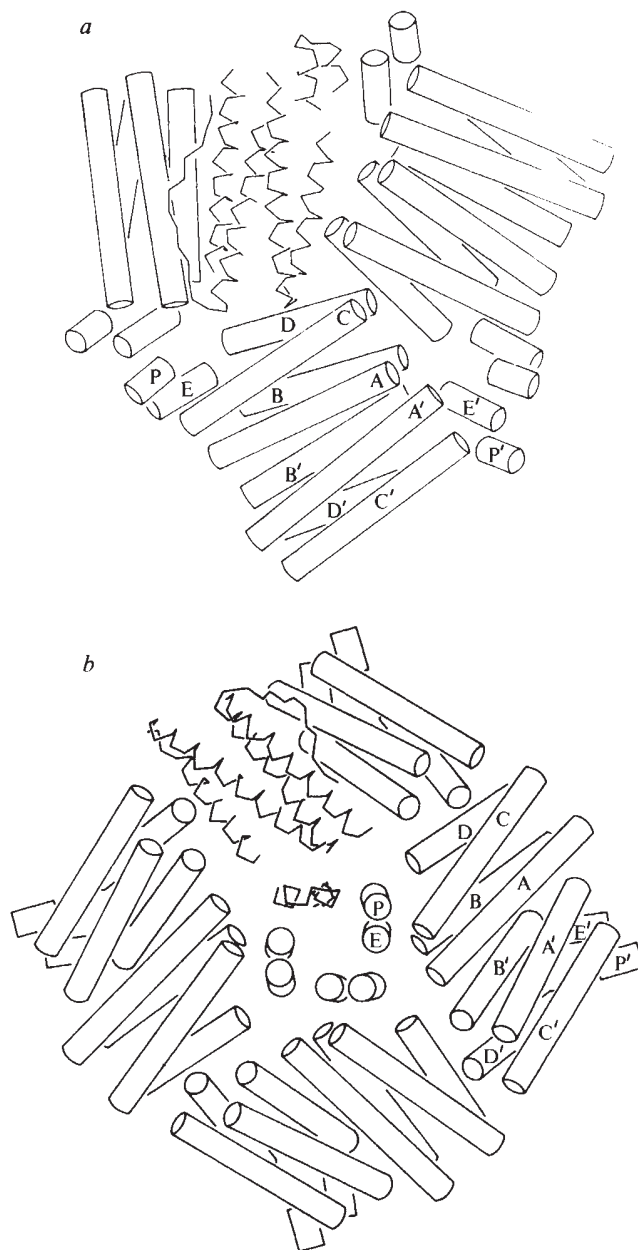


Fig. 2 Illustration of the subunit packing around a molecular triad, showing three dimers (*a*) and around a molecular tetrad, showing four dimers (*b*).

ends of the arrangement. The dimer has a rather flat, lozenge shape, and the overall arrangement within the molecule is such that each dimer lies roughly on the face of a rhombic dodecahedron with subunit axes nearly parallel to one pair of edges of the rhomb. A better molecular envelope is a rhombic dodecahedron truncated to give a polyhedron inscribable within a sphere. Figure 2 shows the packing of three dimers around a three-fold axis (*a*) and four dimers around a four-fold axis (*b*), with α -carbon coordinates used for one subunit only.

Another feature of the structure which has become apparent as a result of model building is a short stretch of anti-parallel β -pleated sheet which joins the loops (L) from neighbouring diad-related subunits over a region around the two-fold axis. Four hydrogen bonds are formed, involving a total of five residues in each subunit, and the β -strands exhibit a characteristic right-handed twist when viewed along their length³. This pleated sheet structure lies on the outside surface of the molecule (see Fig. 1) at a region of intermolecular contact.

The two alternative ways of joining the defined segments which best fit the electron density are as follows. Each of these starts with the same extended N-terminal sequence of about 12 residues preceding the A helix. Only four residues were previously described and a section of electron density, relatively isolated but tenuously connected to these residues, was missed in our first interpretation. The C-terminal end of A may be connected through a short turn to the N-terminus of either B (as in our original interpretation) or C. The peptide segments are joined subsequently in order A $\rightarrow$ B $\rightarrow$ L $\rightarrow$ C $\rightarrow$ D $\rightarrow$ P $\rightarrow$ E (Fig. 3*a*) or A $\rightarrow$ C $\rightarrow$ D $\rightarrow$ P $\rightarrow$ E $\rightarrow$ B $\rightarrow$ L (Fig. 3*b*). Note that all the interhelix connections in both alternative conformations, except for the long loop joining B and C in Fig. 3*a*, exhibit the usual right-handed character⁴ (viewed parallel to the helix-pair contact vector). The alternative subunit conformations are shown schematically in Fig. 3*a* and *b*, respectively, and in *c* and *d* using the mode of representation of Levitt and Chothia⁵. The differences between these connectivities depend on whether density beyond A is attributed to side-chain or main-chain density, and whether the short tail at the C-terminus of E is free, leaving the C-terminus of the molecule inside the shell, or whether it is joined to B. Overall, the chain length seems to be a few residues longer than in our earlier description.

The packing of the four long helices in the apoferritin subunit resembles that of several other proteins, including myo-

Table 1 Angles between the helices within an apoferritin dimer

	A	A'	B'	C'	D'	P'	E'
A		154	29	17	156	142	41
B	171		151	159	20	33	145
C	170	16		174	20	48	129
D	18	168	161		171	138	42
P	62	119	126	51		88	90
E	116	62	57	130	172		94

A, A' and so on are related by the symmetry ($z, -y, x$). Intra-subunit angles, in degrees, are found in the leftmost (lower) triangle and inter-subunit angles on the (upper) right.

haemerythrin^{6,7}, haemerythrin^{8,9}, tobacco mosaic virus¹⁰, cytochrome *b*₅₆₂ (ref. 11) and cytochrome *c'* (P. Weber, personal communication). Moreover, angles between helices shown in Table 1 closely resemble those in myohaemerythrin⁶, and the bundle of four helices in all these structures apparently has a left-handed twist. It does seem that this mode of helical packing may indeed be considered as a stable 'super-secondary structure' as suggested earlier¹². Furthermore, the subunit relationship in the apoferritin shell dimer closely resembles that in cytochrome *c'* (P. Weber, personal communication). An attractive feature of the alternative conformation shown in Fig. 3*b* is that the connectivity closely resembles that in the other proteins. However, a feature of the connectivity of Fig. 3*a*, which is attractive from another point of view, is that this subunit structure seems to be constructed from two approximately equal parts (A to L inclusive and C to E inclusive, comprising sequences of approximately equal length) which could conceivably have arisen from gene duplication. Some support, however slight, for this suggestion is the structural similarity of the A, B and C, D helical pairs and the finding of mercury sites (and therefore presumably cysteine) near the N-terminal ends of helices B and D. An unequivocal assignment of the apoferritin subunit conformation awaits the completion of the amino acid sequence by Crichton and colleagues, and the extension to higher resolution and refinement of the structure. Although it has been reported earlier that the C-terminus is available to carboxypeptidase action, this has not been confirmed in recent work on human liver ferritin (J. M. Sowerby and J. E. Fitton, unpublished preliminary observation). This latter result would favour Fig. 3*a* rather than *b*. Partial proteolysis gives rise to two fragments of molecular weights approximately 7,500 and 11,000, the smaller of which is thought to include the N-terminus^{13,14}. Such fragments could arise in either connectivity by splitting at the BL connection in Fig. 3*a* or at the CD turn in *b*.

We thank the SRC, MRC, the Royal Society and the Blood Research Fund for support, and Mr Graeme Dolderson for technical assistance.

Note added in proof: Details of the cytochrome *c'* structure have now been published¹⁵, and a more detailed comparison of the 4- α -helical proteins has been made¹⁶. Completion of the amino acid sequence of horse spleen apoferritin has also recently been reported¹⁷.

Received 6 June; accepted 26 September 1980.

- Banyard, S. H., Stammers, D. K. & Harrison, P. M. *Nature* **271**, 282-284 (1978).
- Richards, F. M. *J. molec. Biol.* **37**, 225-230 (1968).
- Chothia, C. *J. molec. Biol.* **75**, 295-302 (1973).
- Sternberg, M. J. E. & Thornton, J. M. *J. molec. Biol.* **105**, 367-382 (1976).
- Levitt, M. & Chothia, C. *Nature* **261**, 552-558 (1976).
- Hendrickson, W. A., Klippenstein, G. L. & Ward, K. B. *Proc. natn. Acad. Sci. U.S.A.* **72**, 2160-2164 (1975).
- Hendrickson, W. A. & Ward, K. B. *J. biol. Chem.* **252**, 3012-3018 (1977).
- Ward, K. B., Hendrickson, W. A. & Klippenstein, G. L. *Nature* **257**, 818-821 (1975).
- Stenkamp, R. E., Sieker, L. C., Jensen, L. H. & Loehr, J. S. *J. molec. Biol.* **100**, 23-24 (1976).
- Bloomer, A. C., Champness, J. N., Bricogne, G., Staden, R. & Klug, A. *Nature* **276**, 362-368 (1978).
- Mathews, F. S., Bethge, P. H. & Czerwinski, E. W. *J. biol. Chem.* **254**, 1699-1706 (1979).
- Argos, P., Rossmann, M. G. & Johnson, J. E. *Biochem. biophys. Res. Commun.* **75**, 83-86 (1977).
- Collet-Cassart, D. & Crichton, R. R. in *Proteins of Iron Storage and Transport in Biochemistry and Medicine* (ed. Crichton, R. R.) 185-192 (North-Holland, Amsterdam, 1975).
- Ishitani, K., Niitsu, Y. & Listowsky, I. *J. biol. Chem.* **250**, 3142-3148 (1975).
- Weber, P. C. *et al. Nature* **286**, 302-304 (1980).
- Weber, P. C. & Salemme, F. R. *Nature* **287**, 82-84 (1980).
- Heusterspreute, M., Mathijs, J. M., Wustefeld, C. & Crichton, R. R. *16th Congr. Int. Soc. Hematology*, Montreal (1980).

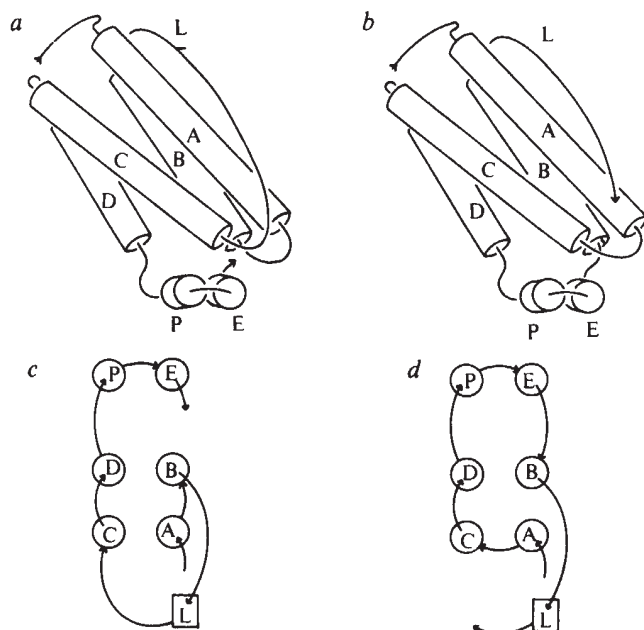


Fig. 3 Two alternative conformations of the apoferritin subunit—A-B-L-C-D-P-E (*a*) and A-C-D-P-E-B-L (*b*)—together with their Levitt-Chothia⁵ representations, *c* and *d*, respectively.

MATTERS ARISING

Evolution of the orang-utan

SMITH AND PILBEAM¹ have argued that "the available evidence...justifies the concept of a Pliocene orang-utan ancestor at least as terrestrial as the modern chimpanzee". It is, however, questionable whether they have established their thesis as equally, or more, plausible than the one they wish to replace (that the ancestral lineage of orang-utans was arboreal).

The authors have used both palaeontological and neontological records, but their arguments are unconvincing. First, there are no unequivocal data on the body sizes of Pleistocene orang-utans so that inferences on arboreality or terrestriality are tenuous. Although Smith and Pilbeam acknowledge the difficulty of drawing conclusions about body size from tooth size alone when considering subfossil orang-utan populations, they conclude that mid-Pleistocene mainland forms were larger-bodied and hence less arboreal than the extant form. The basis for such a conclusion is not evident. Second, plausibility alone does not constitute empirical (or theoretical) support for a thesis. Smith and Pilbeam suggest that it is plausible to view the large body size and marked sexual dimorphism of living orang-utans as remnants of a more terrestrial pattern. This is a possibility but they offer no direct support. Thus there is no reason to consider this alternative of greater importance than any other (sexual selection, niche separation, and such) that may have an influence on whether or not animals are terrestrial. Third, Kay² has shown that (1) there is no relationship between tooth enamel thickness and arboreality or terrestriality, and (2) thick enamel, as found in *Pongo*, seems to be associated with the consumption of hard fruits, nuts and seeds. This weakens arguments by Smith and Pilbeam that ancestral orang-utans may not have been arboreal frugivores.

If there are data that do not support a particular hypothesis, then the inconsistency should be explained. The postcranial anatomy of living orang-utans is the least equivocal source of data on possible "remnants...of a more terrestrial pattern". Morphological studies of orang-utan cheiridia³, wrist⁴, and hip and thigh⁵ have all revealed marked specializations consistent with arboreal progression in a large-bodied primate. No features suggesting adaptations for terrestrial locomotion, present or past, have been reported. Smith and Pilbeam have not accorded these data the attention they deserve, nor explained why extant orang-utans should not display evidence of terrestriality in an ancestral form. It is unclear, for example, whether the authors postulate a stage of terrestriality so brief

that terrestrial adaptations had not time to evolve, or, if arguing for a longer period of terrestriality, whether the return to arboreal progression was to have been accompanied by an exact reversal of evolutionary changes.

Despite the available fossil evidence suggesting that "typical Neogene hominoids (including our hypothetical Pliocene orang-utan) were probably...woodland creatures", one cannot yet eliminate the possibility that this is a simple artefact of the still very limited fossil record. Similarly, there is no reason to expect that all fossils displaying an affinity to extant forms represent populations directly ancestral to living species. It may be worthwhile to consider the possibility, for example, that the mid-Pleistocene orang-utans represented by fossil teeth were a related, but not ancestral, species. Until the competing hypotheses of arboreality or terrestriality can be tested by unequivocal data from the fossil record, they must be considered viable alternatives.

I thank H. McHenry, P. Rodman and L. Scott for reading an earlier version of this manuscript.

L. ALIS TEMERIN

Department of Anthropology,
University of California,
Davis, California 95616

1. Smith, R. J. & Pilbeam, D. R. *Nature* **284**, 447-448 (1980).
2. Kay, R. *Am. J. phys. Anthropol.* **52**, 243 (1980).
3. Tuttle, R. H. *J. Morph.* **128**, 309-364 (1969); *The Chimpanzee* **2**, 167-253 (1970); *Gibbon Siamang* **1**, 136-206 (1972).
4. Lewis, O. J. *Am. J. phys. Anthropol.* **30**, 251-268 (1969); in *Primate Locomotion* (ed. Jenkins, F. A.) 143-169 (Academic, New York, 1974).
5. Sigmon, B. A. *J. Hum. Evol.* **3**, 161-185 (1974).

SMITH AND PILBEAM REPLY—We do not know whether Miocene ancestors of the orang-utan were arboreal or terrestrial, and evidence marshalled in favour of either hypothesis is tenuous at best. Nevertheless, it seems to us that hypotheses for hominoid evolution generally have assumed with confidence that the orang lineage has always been arboreal. In offering evidence for the possibility of a terrestrial ancestor, our purpose was also to lead others to re-evaluate the supposed evidence for arboreality. We completely agree with Temerin's conclusion that arboreality and terrestriality "must be considered viable alternatives".

RICHARD J. SMITH

Department of Orthodontics,
Dental School,
University of Maryland,
Baltimore, Maryland 21201

DAVID R. PILBEAM

Department of Anthropology,
Yale University,
New Haven, Connecticut 06520

Noradrenergic transmission

I WISH to raise two points concerning the discussion and conclusion in a recent report by Hirst and Nield¹.

First, evidence quoted to illustrate that a portion of sympathetic postganglionic transmission is resistant to adrenoreceptor antagonists was derived from work on rodent vasa deferentia: the authors had concluded that in these tissues a component of the motor transmission was 'non-adrenergic' (refs. 2-5). From the reported results¹, however, noradrenergic transmission resistant to α -adrenoreceptor antagonists is postulated. This conclusion may be justified in the case of the arteriolar preparation under study and, if so, constitutes an important observation. However, if this is offered as an explanation for the 'non-adrenergic' component of transmission in vasa deferentia, it should be noted that all previous reports quoted as having supported this latter concept, did so not only on the evidence of resistance to antagonists but also because the response persisted after depletion of noradrenaline by reserpine, for example, to less than 1% of control levels⁶. Furthermore, responses persisted even after chemical sympathectomy when virtually all adrenergic nerve terminals had been removed⁷. In contrast, the 'conventional', adrenergic, α -blocker-susceptible component was completely absent following reserpine or chemical sympathectomy^{2,7}. If allowance for this adrenergic component in the control is made, the 'non-adrenergic' response is not reduced at all following the latter treatments^{2,8}. There is, therefore, no evidence to connect this response in vasa deferentia with noradrenaline or noradrenergic nerves. To associate this with a discussion of the postulated mechanism in arteriolar muscle can only attract irrelevant criticism of the latter. It is also possible that the intracellular excitatory junctional potential and the 'non-adrenergic' contraction of the muscle layers reflect different cellular processes.

Secondly, the conclusion that there are 'two populations of excitatory receptors for noradrenaline in arteriolar smooth muscle' could be confused with the other recent evidence for two distinct types of excitatory α -adrenoreceptor on vascular and non-vascular smooth muscle^{9,10}; in this case, each response was blocked by the appropriate antagonist⁹⁻¹¹. Furthermore, in contrast to the newly reported observation on arteriolar muscle¹, the pressor response to sympathetic nerve stimulation was even more sensitive to blockade by the α_1 -adrenoreceptor antagonist, prazosin, than was the response to exogenous noradrenaline^{11,13}. The 'phentolamine-resistant' effect of

noradrenaline¹ may thus point to a 'third' excitatory mechanism for noradrenaline. This may be supported by the observation that a small, residual, pressor response to catecholamines remains even after combined α - and β -adrenoreceptor blockade¹².

J. C. MCGRATH

*Institute of Physiology,
The University,
Glasgow G12 8QQ, UK*

- Hirst, G. D. S. & Neild, T. O. *Nature* **283**, 767-768 (1980).
- McGrath, J. C. *J. Physiol., Lond.* **283**, 23-39 (1978).
- Ambache, N. & Zar, M. A. *J. Physiol., Lond.* **216**, 359-389 (1971).
- Ambache, N., Dunk, P., Verney, J. & Zar, M. A. *J. Physiol., Lond.* **227**, 433-456 (1972).
- Euler, U. S. von & Hedqvist, P. *Acta physiol. scand.* **93**, 572-573 (1975).
- Gillespie, J. S. & McGrath, J. C. *Br. J. Pharmac.* **52**, 585-590 (1974).
- Booth, F. J., Connell, G. J., Docherty, J. R. & McGrath, J. C. *J. Physiol., Lond.* **280**, 19-20P (1978).
- Docherty, J. R. & McGrath, J. C. *J. Physiol., Lond.* **307**, 18-19 (1980).
- Drew, G. M. & Whiting, S. B. *Br. J. Pharmac.* **67**, 207-215 (1979).
- Docherty, J. R., MacDonald, A. & McGrath, J. C. *Br. J. Pharmac.* **67**, 421P-422P (1979).
- Docherty, J. R. & McGrath, J. C. *Naunyn-Schmiedeberg's Archs Pharmac.* **312**, 107-116 (1980).
- Flavahan, N. A. & McGrath, J. C. *Br. J. Pharmac.* **69**, 355-357 (1980).
- Docherty, J. R. & McGrath, J. C. *Naunyn-Schmiedeberg's Archs Pharmac.* **313**, 101-111 (1980).

HIRST AND NIELD REPLY—We take the view that there are many similarities between the mechanisms of excitatory transmission in arterioles and vasa deferentia of guinea pigs. Both are innervated by catecholamine-containing nerves which when stimulated cause an excitatory junction potential (e.j.p.) in the underlying smooth muscle. The e.j.ps recorded from both tissues are to all intents and purposes identical^{1,2}; e.j.ps from both tissues are rapidly reduced in amplitude by guanethidine and persist in high concentrations of adrenergic blocking agents. We consider that the failure of α -blockade in arterioles arises because the neuronally released transmitter activates receptors which are unlike α -receptors. In our preparation we are able to mimic the action of the neurotransmitter only when noradrenaline is applied to specific regions of the muscle; until similar experiments have been conducted on vas deferentia it seems unnecessary to reject the hypothesis that noradrenaline is the primary transmitter substance³⁻⁶.

McGrath rightly points out that some of the authors who take the view that a second unknown transmitter is released in addition to noradrenaline, base their opinion on further experiments in which tissue noradrenaline was in some way reduced^{3,6}. A common method is by prior treatment with reserpine, which dramatically reduces but does not eliminate tissue noradrenaline⁷. However, it is by no means clear that a dramatic reduction in tissue noradrenaline leads to an equally dramatic reduction in the actual amount of noradrenaline released per nerve impulse⁸.

Concerning McGrath's second point, we did not mean to imply that arteriolar smooth muscle contained two distinct populations of α -receptors, although this may well be the case. We stressed that the 'junctional' receptors that we were able to activate by noradrenaline were completely different from α -receptors. To avoid the possibility of confusion it might perhaps be more appropriate to call the 'junctional' receptors γ gamma-receptors.

G. D. S. HIRST
T. O. NIELD

*Department of Physiology,
Monash University,
Clayton, Victoria,
Australia 3168*

- Hirst, G. D. S. & Neild, T. O. *J. Physiol. Lond.* **280**, 87-104 (1978).
- Byewater, R. A. R. & Taylor, G. S. *J. Physiol., Lond.* **300**, 303-316 (1980).
- Ambache, N. & Zar, M. A. *J. Physiol., Lond.* **216**, 359-389 (1971).
- Ambache, N., Dunk, P., Verney, J. & Zar, M. A. *J. Physiol., Lond.* **227**, 433-456 (1972).
- Euler, U. S. von & Hedqvist, P. *Acta physiol. scand.* **93**, 572-573 (1975).
- Booth, F. J., Connell, G. J., Docherty, J. R. & McGrath, J. C. *J. Physiol., Lond.* **280**, 19P-20P (1978).
- Rand, M. J. & Jurevics, H. *Handbk. Expl. Pharmac.* **39**, 77-159 (1977).
- Enero, M. A. & Langer, S. Z. *Br. J. Pharmac.* **49**, 214-225 (1973).

Germ-line deletion of genes coding for self-determinants

JEMMERSON AND MARGOLIASH's article¹ raises provocative questions about immunological unresponsiveness to self-antigens. They found that the autoantigenic sites on rabbit cytochrome *c* correspond to those regions in the molecule where other mammalian cytochromes *c* differ, and suggested that archaic genes coding for V-region specificities of immunoglobulin recognizing shared portions of the cytochrome *c* molecule were eliminated by evolutionary pressures. This would result in immunologically silent peptide sequences, there remaining mainly V genes for the more recently evolved species-specific determinants. Thus, one mechanism of self-tolerance would be evolutionary germ-line deletion of V genes coding for specificities against self-molecule determinants.

Previous studies by one of us (N.R.R.) were directed to the antigenic properties of thyroglobulin in rabbits². Two sorts of autoantibodies in rabbits were compared, those induced by immunization with rabbit thyroglobulin and those induced by immunization with foreign thyroglobulins, such as those of hog or beef. The first kind of autoantiserum resembled the rabbit antisera to rabbit cytochrome *c* described by Jemmerson and Margoliash. Tested for reaction with rabbit thyroglobulin, the rabbit thyroglobulin readily absorbed all reaction, whereas the hog (or beef) antigen removed only a small part of the antibody content. We concluded that

the rabbit preferred to form antibodies to the non-shared antigenic determinants of thyroglobulin.

In contrast, rabbits immunized with hog thyroglobulin showed relatively weak reactions with rabbit thyroglobulin, and these reactions could be absorbed equally well with hog, beef and rabbit thyroglobulins. Thus, autoantibodies induced by foreign thyroglobulin are directed mainly to the shared antigenic determinants of thyroglobulin.

On the basis of absorption behaviour, then, we could distinguish two types of autoantibodies. Those elicited by thyroglobulin of the same species were directed primarily to species-specific determinants, whereas the autoantibodies evoked by foreign thyroglobulins reacted mainly with the shared antigenic determinants. Of further interest, rabbits immunized with rabbit thyroglobulin showed severe lesions of thyroiditis. Only after repeated injections were even mild lesions seen in rabbits given foreign thyroglobulin.

The basic protein of myelin (BPM), a linear-folded protein without a tertiary structure, is another well characterized autoantigen. There are some 170 amino acid residues, and the sequences for the molecule of several mammalian species reveal a number of substitution sites. Because immunization with BPM induces experimental autoimmune encephalomyelitis (EAE) in all species studied, attention has focused on the capacity of BPM, and various peptides derived from it, to induce disease. The review by Bergstrand³ indicates that different portions of the BPM molecule have greatly differing potency in terms of induction of (1) EAE, (2) cell-mediated immunity and (3) humoral antibody.

In the induction of EAE, which requires the participation of a T-cell response, detailed studies have defined the different amino acid sequences in the homologous proteins which induce EAE in different species. Minor sequence substitutions have an enhancing effect, for example, the encephalitogenic Ser-Thr substitution at residue 79 for the rat⁴, with other changes abrogating activity⁵.

With heterologous BPM injected into rabbits, one of the main determinants of humoral antibody was identified in the region 90-116 (ref. 6), and this does not vary between mammalian species. Unfortunately, relatively little information is available on the amino acid sequence(s) involved in the homologous response. Studies by microcomplement fixation assays in the rabbit⁷ could be interpreted as showing a pattern of cross-reactivity between homologous and heterologous BPM equivalent to that described above for thyroglobulin. In particular, when rabbits were immunized with rabbit BPM, the antibody was predominantly reactive with rabbit BPM but with extensive cross-reaction with BPM of other species. It is assumed, although not

directly tested, that rabbit antibody to BPM is directed to the invariant 92–116 sequence.

Thus for thyroglobulin and BPM, genes specifying reactions to shared determinants do not seem to be eliminated by evolutionary pressures. However, examination of autoantigens in the light of the concept of Jemmerson and Margoliash would be interesting with respect to association of autoimmunity with disease.

IAN R. MACKAY

NOEL R. ROSE

Clinical Research Unit,
The Walter and Eliza Hall
Institute of Medical Research,
Post Office, Royal Melbourne
Hospital,
Victoria 3050, Australia

PATRICK R. CARNEGIE

School of Agriculture,
La Trobe University, Bundoora,
Victoria 3083, Australia

1. Jemmerson, R. & Margoliash, E. *Nature* **282**, 468–471 (1979).
2. Rose, N. R. in *Textbook of Immunopathology* 2 edn Vol. 1 (eds Miescher, P. & Muller-Eberhard, H.) 215–229 (Grune & Stratton, New York, 1977).
3. Bergstrand, H. *Immunochim. Proteins* **1**, 315–370 (1977).
4. Kibler, R. F. *et al. J. exp. Med.* **146**, 1323–1331 (1977).
5. Hashim, G. A. *Immun. Rev.* **39**, 60–107 (1978).
6. Driscoll, B. F., Kramer, A. J. & Kies, M. W. *Science* **184**, 73–75 (1974).
7. Whitaker, J. N. *Comp. Biochem. Physiol.* **59B**, 299–306 (1978).

JEMMERSON AND MARGOLIASH REPLY—A serious problem with the comparative study of protein antigens of unknown amino acid sequence and spatial structure is that one tends to invoke qualitative all-or-none differences, based on the particular antibody-binding test used, when in fact the differences may be merely quantitative. Thus, in the experiments on thyroglobulin cited by Mackay, Rose and Carnegie, it is possible that rabbit thyroglobulin injected into rabbits does not elicit antibodies which bind to hog or beef thyroglobulin because determinant sites on the rabbit protein, which would elicit such cross-reactive antibodies, bind too weakly to the rabbit immune system to effect the needed cell stimulation. In contrast, once antibody production to these same sites has been stimulated by the higher-affinity hog thyroglobulin determinants, the antibodies will bind to the cross-reactive sites on the rabbit protein.

So-called 'shared' determinants can be merely cross-reactive and certainly do not need to be identical. One might be able to distinguish between these categories if the antibodies are tested in assays where binding affinities are measurable (not in immunoadsorption where very low antibody affinities allow binding); in such conditions the appropriate differences in binding have been observed with several

families of protein antigens^{1–4}, as well as with cytochrome *c*. With this last protein family, the data are clearly contrary to any generalization that antibodies reactive with self, elicited by foreign thyroglobulins, react mainly with shared antigenic determinants. Indeed, only one of 13 distinct, site-specific antibody populations isolated from sera of rabbits immunized against foreign cytochromes *c* was shown to bind to a region of the molecule where the host and immunizing proteins were identical⁵. This seems to be the result of a special cross-reaction situation.

With proteins having poorly defined or plastic conformations, a further problem is that amino acid sequence variations distant from a site where amino acid sequences are conserved may possibly change the conformation of that region from the protein of one species to that of another, thus making it antigenic. This is not the case with cytochrome *c* where it has been shown with several different antibody populations that amino acid sequence variations outside the antigenic determinant have no effect on the conformation of that determinant^{5–7}. Moreover, the purification procedure in the case of thyroglobulin seems to change the molecular conformation to some degree⁸, and regions which should be antigenically silent on the basis of amino acid sequence could become immunogenic.

Furthermore, and perhaps most importantly, although our evidence⁹ supports the thesis that evolutionary germ-line deletion of V genes coding for anti-self protein specificities can occur, one need not assume that all selective pressures affecting anti-self antibodies would be negative. Indeed, certain anti-self antibody specificities may be preserved if the antibody itself is important biologically, such as in regulation. Recent observations supporting this idea are that normal animals possess low levels of antibodies having specificities ranging from cellular antigens of the surface¹⁰, cytoplasm¹¹ and organelles (R. J., N. Klinman and E. M., unpublished results) to intermediates in metabolism¹². Anti-idiotypic antibodies which are specific for self-immunoglobulin determinants seem to be involved in regulation of the immune system itself¹³. Certain anti-self antibodies could serve as a clearance mechanism for cell debris resulting from killing of, for example, virus-infected cells, or ageing and death in the normal cell cycle¹⁴. Depending on the nature of the self-antigen (structural lability, physiological location and biological role), the regulatory role of antibodies could vary. Therefore, not all self-antigens would be expected to express only the determinants in evolutionarily variable regions, as was observed with cytochrome *c*. However, immunization with cytochrome *c* has never led to any autoimmune pathology, in sharp contrast to what occurs with thy-

roglobulin and basic myelin protein. Thus, it would be surprising if the V genes corresponding to self-antigenic determinants in the latter two proteins were not as vigorously selected against as those for the former, unless there exists a biological mechanism of which we are unaware.

Finally, even if the V genes corresponding to the antigen studied were treated by the evolutionary process as seems to be the case for cytochrome *c*, one must take into account the appearance of the proteins on the evolutionary time scale relative to the appearance of the antibodies and their average rates of evolutionary change. More recently evolved self-antigens may not show complete antibody deletion of conserved region specificities as not enough time would have been allowed for selective pressures to take effect.

RONALD JEMMERSON

Department of Cellular
and Developmental
Immunology, Scripps Clinic
and Research Foundation,
La Jolla, California 92037

EMANUEL MARGOLIASH

Department of Biochemistry
and Molecular Biology,
Northwestern University,
Evanston, Illinois 60201

1. Atassi, M. Z. in *Immunochemistry of Proteins* Vol 2 (ed. Atassi, M. Z.) 77–176 (Plenum, New York, 1977).
2. Garver, F. A. & Talmage, D. W. *Biochem. Genet.* **13**, 743–755 (1975).
3. Prager, E. M. & Wilson, A. C. *J. biol. Chem.* **246**, 5978–5989 (1971).
4. Welling, G. W., Groen, G., Beintema, J. J., Emmens, M. & Schroder, F. P. *Immunochimistry* **13**, 653–658 (1976).
5. Urbanski, G. J. & Margoliash, E. *J. Immun.* **118**, 1170–1180 (1977).
6. Eng, J. & Reichlin, M. *Molec. Immun.* **16**, 225 (1979).
7. Jemmerson, R. & Margoliash, E. *J. biol. Chem.* **254**, 12706 (1979).
8. Rose, N. R. & Witebsky, E. in *Textbook of Immunopathology* Vol. 1 (eds Miescher, P. & Muller-Eberhard, H.) 150–163 (Grune & Stratton, New York, 1968).
9. Jemmerson, R. & Margoliash, E. *Nature* **282**, 468–471 (1979).
10. Eisenberg, R. A. *et al. J. Immun.* **122**, 2272–2278 (1979).
11. Steele, E. J. & Cunningham, A. J. *Nature* **274**, 483–484 (1978).
12. Bartos, D., Bartos, F., Campbell, R. A., Grettie, D. P. & Smejtek, P. *Science* **208**, 1178–1181 (1980).
13. Jerne, N. K. *Annls. Immun. Inst. Pasteur, Paris* **125c**, 373–389 (1974).
14. Kay, M. M. B. *Proc. natn. Acad. Sci. U.S.A.* **72**, 3521–3525 (1975).

Opiate receptors and adrenal medullary function

THERE is much interest in the role of enkephalins and opiates as transmitters or modulators of neuronal function. Costa and colleagues recently demonstrated¹ that high-affinity opiate receptors are present in the plasma membranes of adrenal chromaffin cells and proposed that "the activation of these receptors causes a non-competitive inhibition of catecholamine release elicited by the stimulation of nicotinic receptors." From

the data presented in their Table 1, it is apparent that morphine and several endogenous opioid peptides inhibit the nicotinic release of catecholamines from isolated chromaffin cells (approximate ID_{50} s: morphine, $>10^{-4}$ M; β -endorphin, 10^{-6} M; Met-enkephalin, $>10^{-4}$ M).

We have performed similar experiments both on freshly isolated bovine chromaffin cells² and on the cells maintained as primary monolayer cultures^{3,4} for up to 16 days. We agree with their findings on inhibition of catecholamine release by morphine (ID_{50} 1.6×10^{-4} M), β -endorphin (5×10^{-5} M), Met-enkephalin (5×10^{-4} M) and Leu-enkephalin ($>5 \times 10^{-4}$ M). Like Costa *et al.*, we found that the nicotine-evoked release but not the K^+ -evoked release of catecholamines was inhibited by these compounds⁵. We also agree that the inhibition by these opiate compounds is noncompetitive with nicotinic agonists; however, we are not convinced by the specificity of the inhibitory effects of opioid compounds for the opiate receptor. In our hands, the two opiate antagonists naloxone (10^{-8} – 10^{-4} M) and naltrexone (10^{-8} – 10^{-4} M) did not reverse the inhibition produced by morphine, but rather they acted like morphine to inhibit the release of catecholamines (naloxone ID_{50} , 10^{-5} M; naltrexone, $>10^{-4}$ M). Further, the effect of naloxone (10^{-6} – 10^{-4} M) and naltrexone (10^{-6} – 10^{-4} M) was additive to that of β -endorphin (5×10^{-5} M) and morphine (10^{-4} M).

We were rather concerned that the levels of morphine and the endogenous opiates required to inhibit catecholamine secretion (10^{-3} – 10^{-5} M) were far in excess of the concentrations of these compounds required for effective receptor occupancy (10^{-8} – 10^{-9} M)^{1,6}. In our search for compounds with higher potency, we found that levorphanol inhibited catecholamine release with an ID_{50} of 4×10^{-6} M. However, dextrorphan, the inactive enantiomer of levorphanol, was equipotent (ID_{50} 4×10^{-6} M) in inhibiting the release of catecholamine from the chromaffin cells.

It is therefore unlikely that the receptor that is involved in these inhibitory actions of morphine and the endogenous opiates on catecholamine release is the same as that described in membrane preparations from the cells¹ and adrenal medullary homogenates⁶ that is stereospecific and has high-affinity binding for Met-enkephalinamide (K_d 0.9×10^{-9} M), naloxone (3.2×10^{-9} M) and naltrexone ($+Na^+$ 1.3×10^{-9} M; $-Na^+$ 3.93×10^{-9} M).

Before any physiological role for opiates in the adrenal gland is considered it is necessary to demonstrate stereospecificity and reversal by antagonists. Although it is possible that slight differences in the way the cells are isolated and maintained in culture could account for the differences we observe, the basic question we raise here of the lack of stereospecificity and the

discrepancy between the levels required for receptor occupancy (10^{-9} M) and the biological effects (10^{-5} – 10^{-4} M) are still relevant and raise serious doubts about the proposed functional role of high-affinity opiate receptors in the modulation of catecholamine release from chromaffin cells.

SIMON LEMAIRE
IRMA LEMAIRE

Department of Pharmacology,
Centre Hospitalier Universitaire,
Sherbrooke, Quebec,
Canada J1H 5N4

DEANNE M. DEAN
BRUCE G. LIVETT

Division of Neurology,
The Montreal General Hospital and
McGill University,
Montreal, Quebec,
Canada H3G 1A4

1. Kumakura, K., Karoum, F., Guidotti, A. & Costa, E. *Nature* **283**, 489–492 (1980).
2. Schneider, A. S., Herz, R. & Rosenheck, K. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5036–5040 (1977).
3. Livett, B. G., Kozousek, V., Mizobe, F. & Dean, D. M. *Nature* **278**, 256–257 (1979).
4. Mizobe, F., Kozousek, V., Dean, D. M. & Livett, B. G. *Brain Res.* **178**, 555–566 (1979).
5. Mizobe, F., Dean, D. M. & Livett, B. G. *Soc. Neurosci. Abstr.* **5**, 534 (1979).
6. Chavkin, C., Cox, B. M. & Goldstein, A. *Molec. Pharmac.* **15**, 751–753 (1979).

COSTA ET AL. REPLY—Lemaire *et al.* have confirmed our report that opioids, when added to primary cultures of chromaffin cells, inhibit the release of catecholamines elicited by nicotinic receptor agonists. They pointed out that this inhibitory effect is nonspecific because the required dose of morphine is too high and the phenomenon lacks stereospecificity. Moreover, as the doses of naloxone required to reverse the action of morphine are also high, at these doses naloxone acts as a partial agonist. Lemaire *et al.* failed to give significance to our finding that lower concentrations of naloxone are required to antagonize the inhibition of catecholamine release caused by β -endorphin. This peptide seems to have a greater affinity and a greater intrinsic activity with the opiate receptor of chromaffin cells which are operative in regulating the action of nicotinic receptor agonists on chromaffin cell stores of catecholamines. In conclusion, the data obtained by us and confirmed by Lemaire *et al.* show that this opiate receptor has peculiar properties (a low sensitivity to morphine, high sensitivity to β -endorphin). Moreover, as expected, the dose of naloxone required to antagonize the inhibition of catecholamine release elicited by opioids depends on the specificity of the agonist used.

These considerations are in line with current thinking suggesting that there are multiple forms of opiate receptors. Hence, it is not surprising that those located on

the membrane of the chromaffin cells have peculiar properties which must be studied before concluding that these receptors do not have a physiological role.

We studied the binding characteristics of the opiate recognition sites of chromaffin cell membranes and compared them with the recognition sites located in synaptic membranes prepared from frontal cortex of bovine brain. The opiate recognition sites of adrenal medulla have a high affinity, and a large capacity to bind etorphine and its structurally related antagonist, diprenorphine. This prompted us to study the action of both compounds on the nicotine-induced release of catecholamines. Diprenorphine, in concentrations of 10^{-8} – 10^{-6} M, is devoid of any action, whereas etorphine prevents the action of nicotine with an IC_{50} of 2×10^{-7} M. The IC_{50} of diprenorphine to inhibit the action of an IC_{50} dose of etorphine is about 3×10^{-7} M. These results show a specific ability of etorphine to antagonize nicotine, and a great specificity of diprenorphine to antagonize etorphine.

We believe that the opiate peptides stored in the splanchnic nerve act on specific receptors located on the chromaffin cell membrane. When these peptides are released from the splanchnic nerve they may function as co-transmitters and modulate the number of receptors for acetylcholine that are available in the chromaffin cell membrane. According to this view, the receptor for the co-transmitter, when occupied by a specific agonist, can influence the number of acetylcholine recognition sites and the extent of the response to acetylcholine.

This type of synaptic regulation could be of interest in the development of new neuropharmacological agents that may modify synaptic transmission by acting on co-transmitter receptors. Theoretically, these potential drugs could minimize or maximize the response to primary transmitters without altering the frequency of neuronal activity, because they will not act on the receptor for the primary transmitter and obliterate the transmitter action. Probably, such drugs would not trigger as many side effects as those that block the primary transmitter receptor. The benzodiazepines are a good example of drugs that act on the co-transmitter regulation. By knowing the nature of various co-transmitters operative in different synapses, we might develop specific drugs that, similarly to the benzodiazepines, will modulate synaptic activity and be relatively devoid of untoward side effects.

E. COSTA
A. GUIDOTTI
L. SAIANI

Laboratory of Preclinical
Pharmacology,
National Institute of Mental Health,
Saint Elizabeths Hospital,
Washington, DC 20032

BOOK REVIEWS

Time for thought

George E. Uhlenbeck

THIS book is a kind of essay about the age-old problem of the nature of time and especially about the question whether time "flows" or whether there is an arrow in time. Since these questions affect every human being, David Park has tried to reach both the general "time amateur" and the scientific expert — a considerable task. I think he has not quite succeeded, but as a result the style of the book is charmingly unpretentious and the author really does try to be clear.

The main line of the book is an attempt to elucidate the development of thought on the two kinds of time, physical time and time of human consciousness, which nowadays seem quite familiar, although they are still full of riddles. In ten short chapters the author traces the development from prehistoric times via the Greek philosophers and the Newtonian and post-Newtonian era to the most recent cosmological theories.

I liked especially the two early chapters entitled "The Greek Questions" and "The Newtonian Answer". The foreshadowing of the two times by Parmenides and Heraclitus, the speculations of Plato about these ideas and the discussion of the atomic theory introduced by Leucippus (perhaps to understand the causal direction in time) I found quite illuminating. The Newtonian chapter starts with a nice evocation of the wonders of the motions in the sky and then the author gives a concise but clear discussion of Newton's laws of dynamics and the law of gravitation. He scrutinizes especially the notion of absolute time, and tries to make clear that in Newtonian dynamics it is *not* necessary to think of a flow of time but that in fact the laws are time-reversible, as we would say nowadays. Here, he also touches on the special theory of relativity, the Einstein-Minkowski space-time concept, the Mach principle and the general theory of relativity. In the appendices he gives more details about these difficult subjects. I am afraid that for the general reader these appendices will be too technical, but they are a good reminder for the physicist, although Appendix 3 on the Mach principle is perhaps rather too speculative.

Now that Newton has clarified the

The Image of Eternity: Roots of Time in the Physical World. By David Park. Pp. 149. (University of Massachusetts Press: 1980.) \$14.50.

universal physical time, the question arises how to measure this time. This is of course done by a clock, but the author asks: What is a clock, and, especially, what is a good clock? He argues that a clock is a device whose law of motion is known, and the better this law is known, the better the clock is. Particularly he discusses the use of the rotation of the Earth from this point of view. This is all quite interesting, but I do have some qualms. Following Ernst Mach, I think that the question of what is a good clock is quite similar to the question of what is a good thermometer. In both cases the answer is essentially determined by convention. From the strictly operational point of view there is in my opinion no "true" time just as there is no "true" temperature.

Park devotes most of the rest of his book to the basic problem of the origin of the idea of the direction or "arrow" in time. If one accepts the Greek concept of the world as a huge collection of moving and interacting atoms, and if one believes that the Newtonian laws of dynamics describe these motions, how can one then reconcile the time-reversibility of these laws with the overwhelming common experience of the natural phenomena? We all know that we grow older and that we have to die. Further, a fundamental law of physics, the second law of thermodynamics, states that a closed system (as the world is supposed to be) tends to a state of equilibrium (or of maximum entropy) in which nothing further happens, the so-called *Wärme Tod*.

Early in this latter part of the book the author gives a short summary of the statistical interpretation of the second law. He illustrates it by a nice discussion of the so-called Maxwell Demon and of the little-known paper by Leo Szilard, which shows why the Demon is not able to violate the second law (in Appendix 1 a simplified but clear summary of Szilard's proof is presented.) Park also emphasizes the enormous length of the so-called Poincaré recurrence time for non-equilibrium states of macroscopic systems, and later argues that as a result these systems can be described causally in time producing our sense of past and future. Hence the arrow

of time on the macroscopic scale is guaranteed by the second law. Although this is certainly the generally accepted view, I feel qualms that the author's presentation now has its faults. I miss, for instance, a deeper discussion of the probabilistic aspects of the second law. No mention is made of the fluctuations around the macroscopic states which produce so many striking phenomena (such as the blue colour of the sky) and which limit the strict causality of the macroscopic laws of motion. But my main reservation is that the author does not make any distinction between living and non-living matter. He says (p. 54) that there is no doubt that our mortality is decreed in some way by the second law. He also does not mention the enormous amount of evidence for the existence of a biological evolution, which seems in such an apparent conflict with the second law. I think this is a serious omission.

The book goes on to tackle the cosmological time problem. After an instructive historical introduction about models of the Universe, the author discusses the modern cosmological "myth", the so-called canonical big bang theory. In my opinion this is the weakest part of the book. It is very speculative and I found it sometimes difficult to follow. The development of the Universe starting from the big bang clearly defines an arrow in time, and I think the author tries to reconcile this with the arrow of time of the second law of thermodynamics. Apparently his arguments have not yet been published in detail. They depend on the notion of microscopic order, which quite escapes me. He touches on the question of a possible end of time for oscillating models of the Universe, and also discusses the possible cosmological significance of the large dimensionless numbers occurring in the present theories.

The last chapter has the charmingly modest title: "Can We Finally Say Anything Sensible?". It is a summary of the basic theme of the book, the existence of two kinds of time, the physical time as measured by clocks and the time of human consciousness which is marked by events. Is there a relation between the two? The author notes that this profound question, which goes back to Plato, is clearly related to the great dilemma between the apparent causality of all natural phenomena and our obvious sense of free will. He makes the

George E. Uhlenbeck is Professor Emeritus of the The Rockefeller University, New York, from where he retired in 1971 as Professor of Theoretical Physics.

interesting suggestion that perhaps the two times are a pair of concepts which are — like the pair of concepts, free will and determinism — complementary in the sense of Bohr. They are two aspects of human knowledge which are both required for our understanding of the world but which exclude each other. Park hopes that this view will lead to a unification of the two times on a higher abstract level. I am

sceptical, mainly because of my qualm about the neglect of the life sciences. To understand time 2 one surely should know what life is!

In all, I think that this is a stimulating book and I gladly recommend it. Each reader will undoubtedly have several objections as I had, but there is also no doubt that the book will make him think again about the roots of time. □

B, Cu, Co, Fe, Mn, Mo, Zn . . .

T.S. West

Applied Soil Trace Elements. Edited by Brian E. Davies. Pp.482. (Wiley: 1980.) £22.50, \$67.50.

PROBLEMS associated with contamination of the environment by traces of heavy metals are of considerable concern to many scientists and laymen. Somewhat strangely, relatively little interest is shown in the equally important problems associated with deficiencies of the trace elements that are essential to all forms of life. Elements such as boron, copper, cobalt, iron, manganese, molybdenum and zinc act as essential catalytic components in most enzyme systems in plants and, boron excepted, in animals where many other elements are vital to the considerably more complex biochemical processes involved.

All trace elements in biological systems are derived from the Earth's crustal rocks originally, but come mainly via weathering of the primary minerals of the parent rocks in soils through plants and, to a lesser extent, through the water supply. This somewhat curiously titled book is a synthesis of most of our knowledge of the trace elements that are known, at the present time, to be "biosignificant". Written by agronomists, chemists and geologists, it is an excellent account which can be recommended to all who are directly concerned with these matters, to others who work in associated areas in the medical sciences and indeed to all who are scientifically aware. It is well written and in many places quite fascinating.

I would have preferred to have seen a fuller account of the chemistry, particularly chemical speciation, of these elements in the soil, and a more systematic consideration of the demands made by plants on the soil. This could have been done at the expense of the rather over-detailed account given of analytical procedures, although one of course does recognize that virtually all present trace-element work in soils is heavily dependent on advances in the analytical sciences.

However, there is no doubt that this book is a major contribution to the subject and in my opinion it is one of the most important to have been published. It is well written and illustrated, and handsomely produced. One hopes that the editor will keep the team together to produce a second edition, when many of the exciting developments that are now taking place — particularly in speciation and synergism between elements — could be given as well as an account of the "new" elements.

This book deserves to be widely read and it probably will. □

Inheritance: the immunological challenge

N.A. Mitchison

Somatic Selection and Adaptive Evolution: On the Inheritance of Acquired Characters. By E. J. Steele. Pp.92. (Williams and Wallace: Toronto/Croom Helm: London, 1980.) \$14.95, £8.95.

THIS rather thin little book advocates the inheritance of acquired characteristics. Its interest lies in the opportunity it offers of observing how a competent young cellular immunologist came to hold such a view, and in the papers published earlier this year by the author, jointly with R. Gorczynski of Toronto, claiming to establish the validity of such inheritance in respect of immunological reactivity. Whether this idea need be considered seriously is a matter of opinion. R. B. Taylor (*Nature* 286, 837; 1980) does so. I do not, on the grounds that such an implausible claim need not be taken seriously to start with except by extreme specialists, in the sense of those willing and able to attempt to repeat the work. Otherwise, let time be the test: if Steele and Gorczynski can make something of their story — pursuing its implications and mechanisms — then let them do so. The rest of us can afford to wait and see. Fortunately, Steele shares the view that their work will stand or fall by what he can accomplish in a follow-up. For example, can heightened as well as diminished reactivity be transmitted by males to their offspring? Can genetically tolerant (heterozygous at H-2) as well as experimentally tolerant males transmit?

The idea is that retroviruses transmit genes from one cell to another by picking up messenger RNA, an old speculation of Temin's. To this is added the thought that the hypothetical flow of genes will favour those genes expressed in clonally expanded cells. As ideas there is nothing wrong with these, except that they are a little old hat now that DNA transfection has become an established technique — which can be used, as M. J. Cline has shown, to transmit methotrexate resistance genes from one mouse to another. The question is whether gene transfer between cells occurs naturally on any appreciable scale. The stability of Mendelian segregation ratios argues that it

does not, particularly when account is taken of segregation from a parent carrying a non-functional allele and from females with X-chromosome inactivation. Steele does not try to rebut or even evaluate this evidence. Instead he quotes a handful of apparent exceptions. Of course, apparent exceptions will turn up if one looks hard enough, but as exceptions they surely deserve critical examination. This they do not receive here.

For instance, Steele cites the apparent transmission of "individual-specific" antibody idiomorph from immunized rabbits to their offspring. But interpreting these data is tricky, because the assignment of idiomorphs to the "individual-specific" as distinct from the "germline-encoded" category is doubtful. Furthermore, when a female is immunized before giving birth, as was the case for the data showing apparent transmission, there is a possibility that passively transmitted antibody may have favoured production of the idiomorph in the offspring. Or to take another instance, Steele cites the apparent production of sickle-cell haemoglobin trait offspring from normal parents as evidence of gene transfer. What is critical here is whether the parents were correctly typed, and to the extent that the typing is more rigorous (that is detects smaller amounts of HbS-type haemoglobin) the argument for gene transfer is weakened. These are technical points, but they indicate the incomplete nature of Steele's argument.

Much the same applies to the musings on the history and future of Lamarckian ideas. A scholarly discussion of these topics would require a coverage of nineteenth-century intellectual history and twentieth-century population genetics which is not the aim here. It is particularly disappointing to find no mention of the problem of conserving genetic variation under any scheme such as that proposed.

If Steele's work can be repeated, and if his interpretation stands, here is one colleague who will most humbly eat his words. □

N.A. Mitchison is Professor of Zoology at University College London.

T.S. West is Director of the Macaulay Institute for Soil Research, Craigiebuckler, Aberdeen.

Anthropology and ecology of the world's savannas

Peter D. Moore

Human Ecology in Savanna Environments. Edited by David R. Harris. Pp.522. (Academic: 1980.) £21.50, \$49.50.

THERE is a romantic appeal about certain of the world's biomes which defies scientific explanation. What is there about desert, rain forest and tundra which excites the imagination, and why does the savanna lack the same magic? It was a similar concern which led David Harris to co-ordinate the Wenner-Gren Foundation conference in 1976 which formed the basis for this book. The general lack of public, and perhaps even scientific, interest in the savanna regions is all the more surprising when one considers the close association of early Man with these tropical grasslands, together with the modern problems of pastoral societies in these regions and the growing conflict with wildlife conservation.

The book is divided into three parts. The first deals with the occupation and exploitation of savanna environments since early prehistoric times. J. Desmond Clark attempts a synthesis of the pollen and vegetation data of Van Zinderen Bakker with the rapidly accumulating information concerning prehistoric communities in Africa. Man, he claims, remained restricted to the open savanna environments until about one million years ago, when *Homo erectus* first invaded the Ethiopian highlands. Hammond provides a similar synthesis for Central and South America, although the palaeoecological data is very much thinner. Within this section of the book also fall a number of anthropological studies of tropical Australia, Africa, India and South-east Asia. To an ecologist, some of the uses of ecological terms in an anthropological context may seem a little misplaced; is it reasonable, for example, to say that pastoral behaviour in East Africa shows the elements of an r-strategy? The statement is largely based upon the process of local resource exhaustion, followed by a moving on to new grazing land. The MacArthur and Wilson model is becoming unrecognizable.

Overall, one is left with the impression, from this first part of the book, that anthropologists have now gathered much field data from their respective study sites and are searching avidly for generalized models.

In the second part of the book, the emphasis moves from savanna history and exploitation systems, to present ecology and management. Here some authors use the ecosystem approach and one is supplied with data concerning production rates and nutrient distributions, particularly in relation to human land use. Among these papers that of P.A. Jewell, concerning the management of game and domestic livestock in African savanna, is outstanding,



East African savanna: Grant's gazelle grazing in the Serengeti.

both in the clarity of its presentation and in the wealth of carefully chosen data it contains. Beside it, the paucity of information from other areas is apparent.

The final part of the book concerns the human biology of savanna peoples. This contains largely physiological, nutritional and epidemiological information concerning the savanna populations, which is vital for an understanding of the ecology of Man in these areas.

One can always think of other subjects which could have been included to make any book more comprehensive in its coverage. High on my list here would be a

contribution by a palaeoecologist, who could put into perspective the limitations of the pollen data, to which so many of the authors refer. Also, a chapter concerning the ecology and importance of fire in savanna would have been helpful. But one must draw the line somewhere, and this book contains more than enough to whet one's appetite and enthusiasm for that neglected biome where Man himself has his roots. □

Peter D. Moore is Senior Lecturer in the Department of Plant Sciences, King's College, London.

Slime mould jubilee

G. Turnock

Growth and Differentiation in Physarum polycephalum. Edited by W. F. Dove and H. P. Rusch. Pp.250. (Princeton University Press: 1980.) £13.70, \$25.

THE true slime mould, *Physarum polycephalum*, has a respected position in the catalogue of lower eukaryotes used for studies of cell biology in general and of differentiation in particular. Two features of its complex life cycle are of especial interest: first, the high degree of synchrony that is characteristic of nuclear division in the coenocytic plasmodium, and, second, the dramatic degree of differentiation involved in the transition from amoebae, which grow and divide as single cells, to plasmodia, which grow without any further cell division. Thus the amoebal-plasmodial transition is a link between two vegetative growth phases; it provides a refreshing contrast for the study of differentiation by comparison with

formation of spores or resting stages — not that *Physarum* cannot supply these as well.

One person, H. P. Rusch, played a particularly important role in stimulating biochemical studies of *Physarum*, and this book celebrates the silver jubilee of the work begun by his group at the McArdle Laboratory at the University of Wisconsin. Cancer research is the remit of the McArdle Laboratory and Rusch chose *Physarum* as a model organism for the basic analysis of growth and differentiation, although he admits in the Introduction that his work on carcinogens was becoming hampered by

James Six

The Construction and Use of a Thermometer by James Six, reviewed in the Autumn Books Supplement of *Nature* (288, 36; 1980), is available from Nimbus Books, 84 Sydney Road, London N10 2RN, price £15 + 60p post and packing.

the development of an allergy to rodents.

With Rusch as one of the co-editors, several of his former collaborators and others have contributed a series of essays describing the progress achieved in the study of *Physarum*. The collection will be of value to all who work with the organism and to others who wish to learn about its peculiarities for the first time.

The potential of genetic studies is emphasized in several of the articles, and the defined strains now available make possible, for example, the biochemical analysis of mutants defective in their ability to undergo the amoebal-plasmodial transition. Heat-sensitive mutants have

been isolated in several laboratories and there is optimism that this class will contribute to our understanding of the control of the synchronous nuclear division cycle in the plasmodial phase. Here, the major challenge, as Holt points out in the first article, is to relate the coordinated replication of the genome to the control of nuclear division.

It is a tribute to the interest that Rusch was able to stimulate that most areas of research on *Physarum* can be covered by his former associates. Transcription, the nuclear replication cycle, genetics and details of the different developmental phases are amongst the topics described.

Throughout there is a discussion of possible areas of progress, although some of the suggestions have a slightly frenetic ring, more appropriate to a grant application than to a collection of critical essays. Another minor criticism is the overlap in details provided in the different chapters. This could have been avoided by stringent editorial control; on the other hand, the argument for repetition is that it does enable each essay to be read individually, and the book as a whole can certainly be recommended. □

G. Turnock is a Senior Lecturer in Biochemistry at the University of Leicester.

All you wanted to know about inflammation

J. L. Gordon

The Cell Biology of Inflammation. Handbook of Inflammation, Vol. 2. Edited by Gerald Weissman. Pp.714. (Elsevier/North-Holland Biomedical: 1980.) Dfl. 230, \$112.25.

THE topics embraced by the title *Cell Biology of Inflammation* cover a fair proportion of many lecture courses in pre-clinical pathology and also touch on several aspects of fundamental research into cell biology — including cell maturation, adhesion, locomotion, phagocytosis, secretion and intercellular interactions —

as well as tissue catabolism, remodelling and repair. Any editor attempting to assemble a book under this catch-all title is faced with a daunting task.

The first point to make, therefore, is that this book succeeds unusually well — better than any other single volume I have read. Several of the authors are acknowledged leaders in their field, and the overall quality is high, with some of the chapters being quite outstanding. I particularly liked Dee Bainton's general view of "The Cells of Inflammation", and there are two excellent contributions on mononuclear

phagocytes from Edelson and from Egal and Davies.

It is impossible to do justice to all the book's 18 chapters in a brief review; the topics cover an astonishingly wide range, from "template activity of isolated lymphocyte nuclei in priming exogenous bacterial RNA polymerase" through "eosinopoiesis and mast cell desensitisation" to "the ultrastructure of pre-lymphatics and tissue channels that constitute the sol phase of interstitial tissue". However, although the coverage is eclectic, this book is not a collection of esoterica — the balance is good and the tone is authoritative, inasmuch as it can be with a subject such as this where so much still remains to be learned.

There are, of course, some disappointments: for example, the reviews of the fast-moving areas of research are already dated, because most of the literature surveys seem to have stopped at the beginning of 1978, and some even earlier. Also, there is inevitable overlap with some parts of Vol. 1 in this series (*Chemical Messengers of the Inflammatory Process*), but this is a much better book than its predecessor — so much so that one wonders what remains for future volumes in this series. The standard of production is high and the copy is clean — Graham Lewis's criticism of the proof-reading of Vol. 1 (*Nature* 286, 89; 1980) perhaps had a salutary effect.

In summary, this is a substantial and important volume, and it deserves to be widely read; it left me with the impression that it could have been subtitled "all you wanted to know about inflammation but were afraid to look for in the original literature". It provides a valuable source of information and reference for those already working on some aspect of the cell biology of inflammation, and it may also stimulate those working on other topics to enter this fascinating field. □

J. L. Gordon is a Principal Scientific Officer in the Department of Cell Biology, ARC Institute of Animal Physiology, Cambridge.

Until winning isn't everything, the public is damned.

Nuclear waste v. cheap energy; the DC-10 disaster v. air travel; genetic engineering v. reducing birth defects. Complex socio-scientific issues like these cannot be adequately addressed by our "either/or" system of dispute resolution, asserts legal scholar Milton R. Wessel. How to truly serve the public interest—by replacing winner-take-all adversary law with "the rule of reason"—is spelled out in *SCIENCE AND CONSCIENCE*, a work that every concerned citizen should read.

**SCIENCE
and
CONSCIENCE**
Milton R. Wessel

\$20.70

COLUMBIA UNIVERSITY PRESS
15A Epsom Road, Guildford, Surrey GU1 3JT